



## King's Research Portal

DOI:

[10.1007/s00453-020-00771-x](https://doi.org/10.1007/s00453-020-00771-x)

*Document Version*

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Ferraioli, D., & Ventre, C. (2020). Approximation Guarantee of OSP Mechanisms: the Case of Machine Scheduling and Facility Location. *ALGORITHMICA*, 0(0), [s00453-020-00771-x]. <https://doi.org/10.1007/s00453-020-00771-x>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Approximation Guarantee of OSP Mechanisms: The Case of Machine Scheduling and Facility Location

Diodato Ferraioli<sup>1</sup> · Carmine Ventre<sup>2</sup>

Received: 28 February 2018 / Accepted: 17 September 2020  
© The Author(s) 2020

## Abstract

Obvious strategyproofness (OSP) is an appealing concept as it allows to maintain incentive compatibility even in the presence of agents that are not fully rational, i.e., those who struggle with contingent reasoning (Li in *Am Econ Rev* 107(11):3257–3287, 2017). However, it has been shown to impose some limitations, e.g., no OSP mechanism can return a stable matching (Ashlagi and Gonczarowski in *J Econ Theory* 177:405–425, 2018). We here deepen the study of the limitations of OSP mechanisms by looking at their approximation guarantees for basic optimization problems paradigmatic of the area, i.e., machine scheduling and facility location. We prove a number of bounds on the approximation guarantee of OSP mechanisms, which show that OSP can come at a significant cost. However, rather surprisingly, we prove that OSP mechanisms can return optimal solutions when they use monitoring—a novel mechanism design paradigm that introduces a mild level of scrutiny on agents’ declarations (Kovács et al. in *WINE* 9470:398–412, 2015).

**Keywords** Mechanism design · Obvious strategyproofness · Approximation ratio · Monitoring

## 1 Introduction

Algorithmic mechanism design (AMD) is by now an established research area in computer science that aims at conceiving algorithms resistant to selfish manipulations. As the number of parties (a.k.a., agents) involved in the computation increases, there is, in

---

A preliminary version of this paper appeared as [13].

---

✉ Diodato Ferraioli  
dferraioli@unisa.it

Carmine Ventre  
carmine.ventre@kcl.ac.uk

<sup>1</sup> DIEM, Università degli Studi di Salerno, Fisciano, Italy

<sup>2</sup> King’s College London, London, UK

fact, the need to realign their individual interests with the designer's. Truthfulness is the chief concept to achieve that: in a truthful mechanism, no selfish and *rational* agent has an interest to misguide the mechanism. A valid question of recent interest is, however, how easy it is for the selfish agents to understand that it is useless (and possibly costly) to attempt to strategize against the truthful mechanism at hand.

Recent research has come up with different approaches to deal with this question. Some authors [1, 5, 9, 39] suggest to focus on “simple” mechanisms; e.g., in posted-price mechanisms one's own bid is immaterial for the price paid to get some goods of interest—this should immediately suggest that trying to play the mechanism is worthless no matter the cognitive abilities of the agents. However, in such a body of work, this property remains unsatisfactorily vague. An orthogonal approach is that of verifiably truthful mechanisms [7], wherein agents can run some algorithm to effectively check that the mechanism is incentive compatible. Nevertheless, these verification algorithms can run for long (i.e., time exponential in the input size) and are so far known only for quite limited scenarios. Importantly, moreover, they seem to transfer the question from the mechanism itself to the verification algorithm.

Li [32] has recently formalized the aforementioned idea of simple mechanisms, by introducing the concept of Obviously Strategy-Proof (OSP) mechanisms. This notion stems from the observation that the practical evidence of truthfulness depends on implementation details. For example, in lab experiments, people facing Vickrey's famous second-price mechanism tend to lie more when this is implemented via a sealed-bid auction than when run via an ascending auction. The quite technical definition of OSP formally captures how implementation details matter by looking at a mechanism as an extensive-form game; roughly speaking, OSP demands that strategy-proofness holds among some gross-grained aggregations of strategy profiles and not only among pairs of strategies profiles (see below for a formal definition). An important validation for the ‘obviousness’ is further provided by Li [32] via a characterization of these mechanisms in terms of agents with limited cognitive abilities (i.e., agents with limited skills in contingent reasoning). Specifically, Li shows that a strategy is obviously dominant if and only if these “limited” agents can recognize it as dominant. OSP is consequently a very appealing notion as in many cases rationality has been seen as the main obstacle to concrete applications of mechanism design paradigms, cf., e.g., Ferraioli et al. [16]; such a relaxation might be a panacea in these cases.

Since its introduction the concept of obviously strategyproofness has been adopted both for allocation problems [3, 6, 18, 32] and for preference aggregation problems [6]. Nevertheless, for all its significant aspects, there appear to be hints that the notion of OSP mechanisms might be too restrictive. Ashlagi and Gonczarowski [3] prove, for example, that no OSP mechanism can return a stable matching—thus implying that the Gale–Shapley matching algorithm is not OSP.

## 1.1 Our Contribution

We investigate the power of OSP mechanisms in more detail from a theoretical computer science perspective. In particular, we want to understand the quality of approximate solutions that can be output by OSP mechanisms. To answer this question, we

focus on two fundamental optimization problems, machine scheduling [2] and facility location [34], arguably (among) the paradigmatic problems in AMD.

For the former problem, we want to compute a schedule of jobs on selfish related machines (i.e., machines with job-independent speeds) so to minimize the makespan. For this single-dimensional problem, it is known that a truthful PTAS is possible [10]. In contrast, we show that there is no better than 2-approximate OSP mechanism for this problem independently from the running time of the mechanism. This result highlights a stark contrast between machine scheduling and auctions. Specifically, [32] characterizes obviously strategyproof mechanisms for every single-dimensional problem where the output is binary (e.g., item won or not); it is not too hard to see that certain optimal algorithms fall within the characterization. Our result, instead, shows that when the output is more general than binary, even simple single-parameter problems become hopeless if obvious strategyproofness is required.

For the facility location problem, we want to determine the location of a facility on the real line given the preferred locations of  $n$  agents. The objective is to minimize the social cost, defined as the sum over the individual agents of the distances between their preferred location and the facility's. Moulin [34] proves that the optimal mechanism, that places the facility on the median of the reported locations, is truthful without money (i.e., the mechanism does not pay or charge the agents). OSP mechanisms without money turn out to be much weaker than that. Badem and Gonczarowski [6] indeed prove that this median mechanism is not OSP. Here, we strengthen this result, by proving in fact a tight bound of  $n - 1$  on the ratio between the social cost output by an OSP mechanism and the optimal social cost. Interestingly, this bound can be shown also for mechanisms that use money, thus showing that transfers are not useful at all to enforce OSP. The proof of this fact uses a novel lower bounding technique for OSP mechanisms wherein the bidding domain (or, equivalently the strategy set) of the lying agent does not necessarily have size two; we in fact show that it is enough to identify two particular values in the bidding domain for our argument to work no matter the size of the domain.

However, a surprising connection of OSP mechanisms with a novel mechanism design paradigm—called *monitoring*—allows us to prove strong positive results. Building upon the notion of mechanisms with verification [26, 35, 36, 42], Kovács et al. [28] introduce the idea that a mechanism can check the declarations of the agents at running time and guarantee that those who overreported their costs end up paying the exaggerated costs. This can be enforced whenever costs can be easily measured and certified. For example, a mechanism can force a machine that in her declaration has augmented her running time to work that long by keeping her idle for the difference between real and reported running time.<sup>1</sup>

<sup>1</sup> Based on the specific setting, this extra cost to misreporting users can be seen as corrective fines or exaggeration penalties (when they involve money) or cost guarantees (when monetary transfers are not allowed). While these terms are more explicit in this context, their general significance is less obvious and ultimately depends on the specific setting at hand. For this reason, in this work we prefer to use the less explicit but more general term “monitoring”, as adopted in previous literature.

We first prove that, no matter the algorithm at hand, there exists an OSP mechanism with monitoring that always compensates the agents their costs. This reimbursement mechanism, introduced in the context of mechanisms with monitoring by Serafino et al. [40] (with the name of *first-price* mechanisms), yields a couple of interesting observations in the context of OSP mechanisms. Firstly, it is the first direct-revelation OSP mechanism. As such it does not need any assumption on the agents' bidding domain nor to repeatedly query/interact with the agents. Secondly, the mechanism uses the algorithm at hand as a black box thus reconciling approximation and obvious strategyproofness. This is relevant because, even for truthful mechanism, it is known that, without monitoring, black box reduction to a non-truthful algorithm may not exist [11].

We remark that similar implementability results are also known for strategy-proofness; in particular, Caragiannis et al. [8] proved that one can design truthful mechanisms for every social function if provided with probabilistic verification and large fines, where probabilistic verification means that each agent has some non-zero probability of being captured lying, and if caught lying she is charged a fine. Our theorem can be then seen as an extension of this result to OSP, where monitoring and reimbursements replace probabilistic verification and large fines (an extension of the results of [8] to OSP that instead consider probabilistic verification is given in [14]. The two sets of assumptions are formally speaking orthogonal to each other. Arguably, however, the designer has much less power in the way she controls the agents in the monitoring setting, and reimbursements are more natural than extravagantly large fees (whose enforceability might be doubtful).

Our general result can be applied to both our problems of interest, and prove the existence of optimal OSP mechanisms with monitoring. Clearly, the optimal mechanism for machine scheduling runs in exponential time; a PTAS that is OSP can however be obtained by plugging in the appropriate approximation algorithm.

Nevertheless, as noted in related literature [37], paying the agents in facility location problems might not be feasible in certain contexts. We therefore look at mechanisms that charge agents to use the facilities (e.g., via a subscription fee); note that, as stated above, transfers ought to be used for good approximations. We design the *interval mechanism* for facility location that is optimal, OSP with monitoring and charges (rather than paying) the agents. This construction adapts the first-price mechanism in [40] to guarantee obvious strategyproofness even in absence of funds to pay the agents.

Our results for facility location draw an interesting parallel between OSP and truthful mechanisms. On the one hand, our bounds for OSP mechanisms without money are in stark contrast with the case of strategyproof mechanisms where the optimum is known to be truthful [34]. On the other hand, our results for OSP

mechanisms can be likened to truthful mechanisms for  $K$ -facility location,  $K \geq 2$ , where there is a linear gap between truthful approximations with and without money [20, 37] (incidentally, there are hints that the gap remains linear also for relaxed notions of truthfulness without money [17])—in the case of OSP mechanisms the price to pay to close the gap is not only money, but also monitoring.

Finally, we focus on the *frugality* [12, 25, 40, 41] of the interval mechanism, i.e., on the possibility to implement an optimal mechanism that is OSP with monitoring and charges agents with less demanding payments. To this aim, we introduce the *optimized interval mechanism (OIM)*, that allows to reduce payments in some cases. We also show that payments assigned by this mechanism cannot be further lowered as long as one insists on OSP direct-revelation mechanisms. Finally we show that even in the most general class of OSP mechanisms, it is impossible to find a mechanism that outperforms OIM on every input.

## 2 Preliminaries

In this work we consider a classical mechanism design setting, in which we have  $n$  selfish agents and a social choice function to be implemented.

Each agent  $i$  has a *true type*  $t_i \in D_i$ , where  $D_i$  is defined as the *domain* of  $i$ . The true type  $t_i$  is *private knowledge* of agent  $i$ . We will denote with  $D = D_1 \times \dots \times D_n$  the set of type profiles. A *social choice function* consists of a *decision policy*  $f$  that associates each type profile  $\mathbf{d} \in D$  with a *public decisions*  $z \in Z$ , and a payment policy  $p$  that assigns to each type profile  $\mathbf{d} \in D$  a *payment profile*  $\mathbf{p} = (p_1, \dots, p_n)$ , that defines for each agent a payment from the mechanism to each agent. Note that the  $p_i$ 's can be positive (meaning that the mechanism will pay the agents) or negative (meaning that the agents will pay the mechanism).

Moreover, each selfish agent  $i$  has a *cost function*  $c_i : D_i \times \mathcal{O} \rightarrow \mathbb{R}$ , where  $\mathcal{O} = Z \times \mathbb{R}^n$ . For  $d_i \in D_i$  and  $O \in \mathcal{O}$ ,  $c_i(d_i, O)$  is the cost paid by agent  $i$  to implement  $O$  when her type is  $d_i$ . Note that we presented our setting for agents having a cost to implement the solution chosen by the mechanism; clearly, in some settings, as for example auctions, agents have a non-negative valuation (i.e., non-positive cost) for the outcome. Our positive results extend also to these settings, as we do not rely on the sign of the costs.

**Mechanism Implementation Tree** A mechanism then consists of a protocol that produces an *outcome*  $(z, \mathbf{p}) \in \mathcal{O}$ . To this aim, the mechanism is allowed to interact with agents. During this interaction, agent  $i$  is observed to take *actions* (e.g., saying yes/no). Each action of agent  $i$  may signal some properties that only a subset of the types in  $D_i$  enjoy (e.g., in English auctions saying “yes” signals that the valuation of the agent for the item is above the current price, and saying “no” signals that it is below). In other words, each action of  $i$  may restrict the set of types of  $i$  to a subset of  $D_i$ . To stress this, we then say that agent  $i$  takes an *action according to*  $d_i$ , where  $d_i$  is any of the types of  $i$  as signaled by the actions taken by this agent. Clearly,

the choice of the action taken by the agent can depend on the current status of the mechanism's execution.

We formally model this through the concept of mechanism implementation tree. Our definition is different from the original one given in Li [32], and is based on the concept of *round-table mechanism* defined in Mackenzie [33]. (The two definitions are formally proved to be equivalent for what concerns the definition of obvious strategyproof mechanisms.) Specifically, a *mechanism implementation tree* is a directed tree  $T = (V, E)$  such that:

- every leaf  $\ell$  of the tree is labeled by a possible outcome  $O(\ell) \in \mathcal{O}$  of the mechanism;
- every internal vertex  $u \in V$  is labeled by an agent  $S(u) \in [n]$  of agents;
- every edge  $e = (u, v) \in E$  is labeled by a subset  $D(e) \subseteq D$  of type profiles such that:
  - the subsets of profiles that label the edges outgoing from the same vertex  $u$  are disjoint, i.e., for every triple of vertices  $u, v, v'$  such that  $(u, v) \in E$  and  $(u, v') \in E$ , we have that  $D(u, v) \cap D(u, v') = \emptyset$ ;
  - the union of the subsets of profiles that label the edges outgoing from a non-root vertex  $u$  is equal to the subset of profiles that label the edge going in  $u$ , i.e.,  $\bigcup_{v: (u, v) \in E} D(u, v) = D(\phi(u), u)$ , where  $\phi(u)$  is the parent of  $u$  in  $T$ ;
  - the union of the subsets of profiles that label the edges outgoing from the root vertex  $r$  is equal to the set of all profiles, i.e.,  $\bigcup_{v: (r, v) \in E} D(r, v) = D$ ;
  - for every  $u, v$  such that  $(u, v) \in E$  and for every two profiles  $\mathbf{d}, \mathbf{d}' \in D(\phi(u), u)$  such that  $d_i = d'_i$ , where  $i = S(u)$ , if  $\mathbf{d}$  belongs to  $D(u, v)$ , then also  $\mathbf{d}'$  must belong to  $D(u, v)$ .

Roughly speaking, the tree represents the steps of the execution of the mechanism. As long as the current visited vertex  $u$  is not a leaf, the mechanism interacts with agent  $S(u)$ . Different edges outgoing from vertex  $u$  are used for modeling the different actions that this agent can take during this interaction with the mechanism. As stated above each action signals some property that only some subset of types of  $S(u)$  can satisfy. Hence, by taking a specific action at node  $u$ ,  $S(u)$  is publicly announcing that her type belongs to the subset corresponding to that action. Our definition allows to identify the publicly announced subset of types with each action through the label  $D(e)$  on edge  $e = (u, v)$ : this lists all type profiles in  $D(\phi(u), u)$  in which the type of  $S(u)$  belongs to the subset of types corresponding to the action associated to edge  $e$ . Note that we can safely assume that different actions must signal different subset of types (indeed, if two different actions are signalling the same types we can consider them as a single action). The execution ends when we reach a leaf  $\ell$  of the tree. In this case, the mechanism returns the outcome that labels  $\ell$ .

Observe that, according to the definition above, for every profile  $\mathbf{d}$  there is only one leaf  $\ell = \ell(\mathbf{d})$  such that  $\mathbf{d}$  belongs to  $D(\phi(\ell), \ell)$ . For this reason we slightly abuse notation, and say that  $\mathcal{M}(\mathbf{d}) = O(\ell)$ , where  $\mathcal{M}(\mathbf{d})$  denotes the outcome returned by the mechanism when agents take actions according to types  $\mathbf{d} = (d_1, \dots, d_n)$ . The mechanism is then said to *implement* a social choice function

$(f, p)$  if  $\mathcal{M}(\mathbf{d}) = (f(\mathbf{d}), p(\mathbf{d}))$  for every type profile  $\mathbf{d} \in D$ . A mechanism is said *without money* if the social choice function that it implements sets  $p(\mathbf{d}) = 0^n$ , i.e., it assigns a null payment to every agent  $i$  and for every profile  $\mathbf{d} \in D$ . Moreover, a mechanism  $\mathcal{M}$  is said to satisfy *voluntary participation* (aka, *individual rationality*) if for every  $i$  and every  $\mathbf{d}_{-i} = (d_1, \dots, d_{i-1}, d_{i+1}, \dots, d_n)$ , it holds that  $c_i(t_i, \mathcal{M}(t_i, \mathbf{d}_{-i})) \leq 0$ , where  $t_i$  is the true type of  $i$ .

**Obvious Strategyproofness** Let us now formally define the concept of obviously strategyproof mechanism. This concept has been introduced in [32]. However, we here provide a different definition based on [33].

Given a mechanism  $\mathcal{M}$  with implementation tree  $\mathcal{T} = (V, E)$ , for every type profile  $\mathbf{d}$  and every node  $u \in V$ , we say that  $\mathbf{d}$  is *compatible* with  $u$  if  $\mathbf{d} \in D(\phi(u), u)$ . Finally, two profiles  $\mathbf{d}, \mathbf{d}'$  are said to *diverge* at vertex  $u$  if there are two distinct vertices  $v, v'$  such that  $(u, v) \in E, (u, v') \in E$  and  $\mathbf{d} \in D(u, v)$ , whereas  $\mathbf{d}' \in D(u, v')$ .

An  $\mathcal{M}$  with implementation tree  $\mathcal{T}$  is *obviously strategyproof* (OSP) if for every agent  $i$ , for every vertex  $u$  such that  $i = S(u)$ , for every  $\mathbf{d}_{-i}, \mathbf{d}'_{-i}$  (not necessarily different), and for every  $d_i \in D_i$ , such that  $d_i$  is different from the true type  $t_i$  of  $i$ , and both  $(t_i, \mathbf{d}_{-i})$  and  $(d_i, \mathbf{d}'_{-i})$  are compatible with  $u$ , but diverge at  $u$ , it holds that  $c_i(t_i, \mathcal{M}(t_i, \mathbf{d}_{-i})) \leq c_i(t_i, \mathcal{M}(d_i, \mathbf{d}'_{-i}))$ . Roughly speaking, an obvious strategyproof mechanism requires that, at each time step agent  $i$  is asked to take a decision that depends on her type, the worst cost that she can pay if at this time step she behaves according to her true type is at least the same as the best cost achievable by different strategies. Note, in fact, how the definition considers inconsistent strategies, where agents take actions according to different types at different nodes of the tree. In fact, for each node  $u$  of the tree in which player  $i$  can separate  $t_i$  from a different type in  $D_i$  with her actions, we compare her cost for each of the leaves below  $u$  in which she plays according to her type  $t_i$  with *all* of the leaves below  $u$  where she deviates to something different from  $t_i$ . (We remark that our definition of  $c_i$  takes the payments into account.)

This also highlights the difference with *strategyproof* mechanisms  $\mathcal{M}$ : in a strategyproof mechanism always taking actions according to the true type is dominant for each agent; then in strategyproof mechanisms we do not compare all the leaves below a tree node  $u$  where agent  $i$  can separate her true type from other reports in her domains but we only focus on the leaves where  $\mathbf{d}_{-i}$  is the same. Hence, if a mechanism is obviously strategyproof, then it is also strategyproof. Indeed, the latter requires that an agent would be able to consider all profiles of other agents' types and to evaluate for each of these that truthful behavior provides a larger welfare. In obviously strategyproof mechanisms, an agent only needs to look at aggregates of the other agents' types profiles, roughly the worst type profile and the best one, and it is required to evaluate that truthful behavior is the best option only against these two aggregates. Clearly, being the most profitable choice even against the worst type profile implies that truthful behavior is the best option for every profile, and thus strategyproofness.



However, the opposite does not necessarily hold. The simplest example of this fact is given by sealed-bid second-price auctions that are known to be strategyproof and to produce the same outcome as English ascending-price auctions. However, for the latter it can be easily proved they are OSP, whereas for the former this is not the case [32].

It is important in our context to focus on mechanisms that are not trivial. We say that a mechanism with implementation tree  $\mathcal{T}$  is *trivial* if for every vertex  $u \in V$  and for every two type profiles  $\mathbf{d}, \mathbf{d}'$ , it holds that  $\mathbf{d}$  and  $\mathbf{d}'$  do *not* diverge at  $u$ . That is, a mechanism is trivial if it never requires that agents take actions that depend on their type. Observe that if a mechanism  $\mathcal{M}$  is not trivial, then every path from the root to one leaf goes through a vertex  $u^*$  such that there are two type profiles  $\mathbf{d}, \mathbf{d}'$  that diverge at  $u^*$ . Since  $\mathbf{d} \neq \mathbf{d}'$ , then there exists agent  $i^* = S(u^*)$  such that  $d_{i^*} \neq d'_{i^*}$ . For this reason, we call  $i^*$  the *divergent agent* for the mechanism  $\mathcal{M}$ . Note that the divergent agent takes a decision that depends on her own type before any other agents revealed any information about their own type. For this reason, in order to prove that a mechanism is not obviously strategyproof, it is sufficient to show that there are two type profiles  $\mathbf{d}, \mathbf{d}'$  with  $d_{i^*} \neq d'_{i^*}$  such that they diverge at  $u^*$ , and  $c_{i^*}(d_{i^*}, \mathcal{M}(\mathbf{d})) > c_{i^*}(d_{i^*}, \mathcal{M}(\mathbf{d}'))$ .

Let us state two further properties of obvious strategyproofness, that turn out to be very useful in the rest of the paper. First, if  $\mathcal{M}$  is OSP when the type profile is taken from  $D$ , then it continues to be OSP even if the types are only allowed to be selected from  $D' = D'_1 \times \dots \times D'_n$ , where  $D'_i \subseteq D_i$  [32]. Moreover, let us define  $\mathcal{M}'$  obtained from  $\mathcal{M}$  by *pruning*, i.e. removing, the paths involving actions corresponding to types in  $D \setminus D'$ . If  $\mathcal{M}$  is OSP, then also  $\mathcal{M}'$  enjoys this property [32].

**Direct Revelation Mechanisms**  $\mathcal{M}$  is said to be a *direct-revelation* mechanism if, for every  $i$ , the set of actions that agent  $i$  can take during the interaction with the mechanism coincides with her domain  $D_i$  and its (obvious) dominant strategy is to always take the action that coincides with her true type  $t_i$ .

Observe that direct-revelation mechanisms are not constrained in the order in which agents interact with mechanism (in particular, in order to fit our definition of OSP mechanisms, we will assume that agents reveal their types sequentially).

**Monitoring** Commonly, the cost incurred by agent  $i$  for the outcome  $(z, \mathbf{p})$  is defined as a quasi-linear combination of the *expense* faced by an agent of type  $t_i$  for implementing the public decision  $z$ , that with a slight abuse of notation we denote as  $t_i(z)$ , and the payment  $p_i$ , i.e.,  $c_i(t_i, (z, \mathbf{p})) = t_i(z) - p_i$ . This approach disregards the agent's declaration for evaluating her cost.

In mechanisms with monitoring the usual quasi-linear definition is maintained but costs paid by the agents are more strictly tied to their declarations [28]. Specifically, in a mechanism with monitoring  $\mathcal{M}$ , if  $z$  is the public decision returned by the mechanism when agents behave as if their types was as in the profile  $\mathbf{d}$ , then the expense that  $i$  would incur if her true type was the declared type  $d_i$  is considered a lower bound on the real agent  $i$ 's expense for  $z$ . So an agent is allowed to have a real expense higher than  $d_i(z)$  but not lower.<sup>2</sup> Formally, we have<sup>3</sup>

$$c_i(t_i, \mathcal{M}(\mathbf{d})) = \max\{t_i(f(\mathbf{d})), d_i(f(\mathbf{d}))\} - p_i(\mathbf{d}).$$

Note that practical implementation is possible whenever evidences of the expense faced by an agent can be provided (and cannot be counterfeited). Indeed, once the evidence is provided, the mechanism can check whether  $t_i(f(\mathbf{d})) < d_i(f(\mathbf{d}))$  and charge the agent the difference for cheating.<sup>4</sup>

We next describe two specific problems of interest.

**Machine Scheduling** Here, we are given a set of  $m$  different jobs to execute and the  $n$  agents control related machines. That is, agent  $i$  has a job-independent processing time  $t_i$  per unit of job (equivalently, an execution speed  $1/t_i$  that is independent from the actual jobs). The set of public decisions  $Z$  consists of all possible schedules  $\mathbf{z} = (z_1, \dots, z_n)$  of jobs to the machines, where  $z_i \in \{0, \dots, m\}$  and  $\sum_i z_i = m$ . Here  $z_i$  denotes the job load assigned to machine  $i$ . The expense that agent  $i$  faces for implementing the schedule  $\mathbf{z}$  is  $t_i(\mathbf{z}) = t_i \cdot z_i$ . Note that our mechanisms for machine scheduling will always pay the agents.

We focus on mechanism implementing social choice functions whose decision policy  $f^*$  optimizes the *makespan*, i.e.,

$$f^*(\mathbf{d}) \in \arg \min_{\mathbf{z}} mc(\mathbf{z}, \mathbf{d}), \quad mc(\mathbf{z}, \mathbf{d}) = \max_{i=1}^n d_i(\mathbf{z}).$$

<sup>2</sup> We highlight that the designer only checks that agents are not “better” than declared. That is, agents can pretend to have a higher expense but this will force him to pay at least that amount. Agents can still underbid and at execution time have a higher expense (e.g., they could say to have underestimated their expenses). Note that contrarily to the notion of verification in [35] there is here no punishment for this misbehavior.

<sup>3</sup> Observe that this definition is tailored to direct revelation mechanisms, where each strategy profile is associated to a different leaf of the implementation tree. Since our positive result uses direct revelation mechanisms, this is not an issue. For general extensive-form mechanisms, each leaf can correspond to more than one profile. The definition of monitoring should then specify which type  $d_i$  compatible with the leaf reached ought to be used to monitor agent  $i$ —see discussion in [30]. For our negative results, we focus on a pessimistic notion of monitoring where  $c_i(t_i, \mathcal{M}(\mathbf{d})) = \max\{t_i(f(\mathbf{d})), \min_{d'_i \text{ compatible with } \ell(\mathbf{d})} d'_i(f(\mathbf{d}))\} - p_i(\mathbf{d})$ , where  $\ell(\mathbf{d})$  denotes the unique leaf associated to the profile  $\mathbf{d}$ .

<sup>4</sup> One relevant applicative scenario here is for example for reimbursement of previously declared expenses. These expenses are usually reimbursed only upon production of receipts so that for agents to be consistent with overbidding they need to pay the exaggerated (reported) cost. Receipts are in this case the evidence that one needs to “monitor” the agents.

We say that the mechanism  $\mathcal{M}$  is  $\alpha$ -approximate if the public decision  $z$  returned by the mechanism when agents behave according  $\mathbf{d}$  has a makespan that is at most a factor  $\alpha$  away from decision returned by  $f^*$  on the same input. In this case, we also say that  $\alpha$  is the approximation ratio of mechanism.

In this setting, monitoring means that those agents who have exaggerated their unitary processing time, i.e., they take actions according to  $d_i > t_i$ , can be made to process up to time  $d_i$  instead of the true processing time  $t_i$ . Note that this can be easily achieved in many real-world scenarios even without “evidence”, by simply programming jobs so that they run dummy operations until time  $d_i$  if the outcome is returned before such time.

**Facility Location** In the facility location problem, the type  $t_i$  of each agent consists of her position on the real line. The set of public decisions  $Z$  consists of a position  $z \in \mathbb{R}$  for the facility. The expense that agent  $i$  pays for having the facility in position  $z$  is  $t_i(z) = d(t_i, z) = |t_i - z|$ . So,  $t_i(z)$  denotes the distance between  $t_i$  and the location of the facility.

We focus on mechanisms implementing social choice functions whose decision policy  $f^*$  optimizes the *social cost*, i.e.,

$$f^*(\mathbf{d}) \in \arg \min_{z \in \mathbb{R}} \text{cost}(z, \mathbf{d}), \quad \text{cost}(z, \mathbf{d}) = \sum_{i=1}^n d_i(z).$$

As above, we say that the mechanism  $\mathcal{M}$  is  $\alpha$ -approximate if the public decision  $z$  returned by the mechanism when agents behave according  $\mathbf{d}$  has a social cost that is at most a factor  $\alpha$  away from decision returned by  $f^*$  on the same input, and we denote  $\alpha$  as the approximation ratio of the mechanism.

### 3 A General Positive Result

For an decision policy  $f$ , define the payment policy  $p$  such that  $p_i(\mathbf{d}) = d_i(f(\mathbf{d}))$ . We call the direct-revelation mechanism (that implements)  $(f, p)$  a *reimbursement mechanism*.

**Theorem 1** *Any direct-revelation reimbursement mechanism is OSP with monitoring and satisfies voluntary participation, whatever the order in which agents interact with the mechanism.*

**Proof** In order to prove that  $\mathcal{M} = (f, p)$  is OSP, consider agent  $i$  and let  $t_i$  be her true type. We next show that for agent  $i$ , being truthful always minimizes her cost, regardless of the decisions taken by other agents. To this aim, let us recall that in a mechanism with monitoring the cost that  $i$  pays, given the submitted type profile is  $\mathbf{d}$ , is

$$c_i(t_i, \mathcal{M}(\mathbf{d})) = \max\{t_i(f_i(\mathbf{d})), d_i(f_i(\mathbf{d}))\} - p_i(\mathbf{d}).$$

Suppose that  $i$  is truthful; then for every  $\mathbf{d}_{-i}$ , it turns out that

$$c_i(t_i, \mathcal{M}(t_i, \mathbf{d}_{-i})) = t_i(f_i(t_i, \mathbf{d}_{-i})) - p_i(t_i, \mathbf{d}_{-i}) = 0.$$

Suppose, instead, that  $i$  lies and says  $d_i$ . Then for all  $\mathbf{d}_{-i}$ , if  $d_i(f_i(\mathbf{d})) > t_i(f_i(\mathbf{d}))$  then

$$c_i(t_i, \mathcal{M}(d_i, \mathbf{d}_{-i})) = d_i(f_i(d_i, \mathbf{d}_{-i})) - p_i(d_i, \mathbf{d}_{-i}) = 0;$$

if, instead,  $t_i(f_i(\mathbf{d})) \geq d_i(f_i(\mathbf{d}))$  then

$$c_i(t_i, \mathcal{M}(d_i, \mathbf{d}_{-i})) = t_i(f_i(d_i, \mathbf{d}_{-i})) - p_i(d_i, \mathbf{d}_{-i}) > 0.$$

Note that all the cost (in)equalities hold no matter the value of  $\mathbf{d}_{-i}$ . Thus, in both cases the best cost that  $i$  can obtain by adopting a strategy different from the truthful one is not smaller than the worst cost that  $i$  can obtain by adopting the truthful strategy, as desired.  $\square$

It is important to note that in the construction above, we may use every algorithm as a *black box*. This in particular means that we can turn any optimal (approximation, resp.) algorithm into an optimal (approximate, resp.) OSP mechanism with monitoring (without losses to the approximation guarantee, resp.). Thus, for Combinatorial Auctions (CAs) with additive bidders our mechanism with monitoring beats the lower bound proved by Badem and Gonczarowski [6] for OSP mechanisms. Just as the weaker notion of verification has been shown to be useful in the context of truthful CAs without money [21], our result shows that OSP with monitoring matches the best-known (polynomial-time) approximations achieved not only by truthful mechanisms [29, 31], but also by general algorithms [22, 24].

We also stress that this construction is *query optimal*, as the interaction with each agent minimizes the number of queries. It is worthy to observe that such an interaction does not need to be simultaneous since obvious strategyproofness is maintained even if agents are queried in an adversarially chosen order and know what the bidders preceding them have declared. We will see how to exploit this property to reduce our payments for facility location. Finally, as observed above, we do not require the domain of each agent to be finite.

It should be noted that the reimbursement mechanism might require very large payments. Just like frugality is an important research agenda for strategyproof mechanisms [12, 25], our result paves the way for further research that looks at obviously strategyproof mechanisms, especially indirect revelation ones, with lower payments. In Sect. 5.2, we will give a first sample of this approach for the facility location problem.

## 4 Machine Scheduling

We now show that, without monitoring, there is no OSP mechanism that satisfies voluntary participation and returns an assignment of jobs to machines whose makespan is less than twice the makespan of the optimal assignment. Hence, while in binary allocation problems, optimal OSP mechanisms are possible [6, 32] only within a single parameter setting, for scheduling optimality turns out to be impossible even in this setting.

Interestingly, this is the same lower bound that Nisan and Ronen [35] proved for the approximation ratio of strategyproof mechanisms for *unrelated* machines, i.e., when it is not possible to express the processing time of jobs on machines as a product of jobs' load and machine's unit processing time. We wonder if one can improve the lower bound for OSP mechanisms with related machines in order to match the best known lower bounds for SP mechanisms with unrelated machines, i.e.,  $1 + \phi \approx 2.61$  for general mechanisms [27], and  $n$  for anonymous mechanisms [4].

**Theorem 2** *For every  $\varepsilon > 0$ , there is no  $(2 - \varepsilon)$ -approximate mechanism for the scheduling problem on related machines that is OSP without monitoring and satisfies voluntary participation.*

**Proof** Let us consider the simple setting in which there are exactly two machines, that we denote with 0 and 1, and two equivalent jobs of unit length. We will denote with  $t_0$  and  $t_1$  the type, i.e., the job processing time, of machine 0 and 1, respectively. Suppose there is a  $k$ -approximate, with  $k < 2$ , OSP mechanism  $\mathcal{M}$  that satisfies voluntary participation.

Since the mechanism is  $k$ -approximate, then it must be the case that: if  $t_0 < \frac{t_1}{2k}$ , then  $\mathcal{M}$  assigns both jobs to machine 0; if  $t_0 > 2k \cdot t_1$ , then  $\mathcal{M}$  assigns both jobs to machine 1; if  $\frac{k}{2} \cdot t_1 < t_0 < \frac{2}{k} \cdot t_1$ , then  $\mathcal{M}$  assigns one job to each machine.

Moreover, since mechanism  $\mathcal{M} = (f, \mathbf{p})$  is OSP, then it must be also strategyproof. Archer and Tardos [2] proved that a mechanism for the machine scheduling problem is strategyproof and satisfies voluntary participation if and only if (i) the allocation of jobs to machine  $i \in \{0, 1\}$  returned by  $f$  when the type of the other machine is  $t_{1-i}$  is *monotone*, i.e.,  $f_i(t_i, t_{1-i}) \leq f_i(t'_i, t_{1-i})$  whenever  $t_i > t'_i$ ; (ii) the payment that the machine  $i$  receives is

$$p_i(t_i, t_{1-i}) = t_i f_i(t_i, t_{1-i}) + \int_{t_i}^{\infty} f_i(x, t_{1-i}) dx.$$

In our setting, the monotonicity requirement implies that, for every  $t_{1-i}$ , there are  $t' \in \left[\frac{t_{1-i}}{2k}, \frac{k}{2} \cdot t_{1-i}\right]$  and  $t'' \in \left[\frac{2}{k} \cdot t_{1-i}, 2k \cdot t_{1-i}\right]$ , such that machine  $i$  is assigned both jobs if  $t_i < t'$ , only one job if  $t' \leq t_i \leq t''$ , and no jobs if  $t_i > t''$ . Hence,  $p_i(t_i, t_{1-i}) = t' + t''$  if  $t_i < t'$ ,  $p_i(t_i, t_{1-i}) = t''$  if  $t' \leq t_i \leq t''$ , and  $p_i(t_i, t_{1-i}) = 0$  otherwise.

Let us now restrict the domain of the agents to  $D' = \{a, b\}^2$ , with  $b > k^2 a$ . Let  $\mathcal{M}'$  be the mechanism obtained by pruning  $\mathcal{M}$  according to this restriction. As

stated above,  $\mathcal{M}'$  must be an OSP mechanism. Moreover, the approximation ratio of  $\mathcal{M}'$  cannot be worse than the approximation ratio of  $\mathcal{M}$ . Hence,  $\mathcal{M}'$  cannot be trivial (indeed, a trivial mechanism would have approximation ratio worse than  $k$ ).

Let  $i$  be the divergent agent of  $\mathcal{M}'$ . Clearly,  $a$  and  $b$  are the types in which  $i$  diverges. Suppose that  $t_i = a$ . If  $i$  behaves according to  $t_i$ , then it may be the case that the other agent behaves according type  $a$  too. As showed above, in this case machine  $i$  receives one job and payment  $t'' \leq 2ka$ . Hence,  $c_i(a, \mathcal{M}(a, a)) \geq a - 2ka$ . Suppose instead that  $i$  behaves as if her type was  $b$ . It may be the case that the other agent behaves according type  $b$  too. Then, machine  $i$  still receives one job and a payment  $t'' \geq \frac{2}{k} \cdot b$ . Hence,

$$c_i(a, \mathcal{M}(b, b)) \leq a - \frac{2}{k} \cdot b < a - 2ka = c_i(a, \mathcal{M}(a, a)),$$

where we used that  $b > k^2a$ . In words, the best cost paid by  $i$  if she does not behave according to her true type can be lower than the worst cost she can pay if she behaves according to her true type. Then, the mechanism  $\mathcal{M}'$  is not OSP, contradicting our hypothesis.  $\square$

Since there is a PTAS for the allocation of jobs to related machines [23], then we have the following corollary of Theorem 1.

**Corollary 1** *There is an OSP mechanism with monitoring that computes the optimal scheduling of jobs to related machines (in exponential time). Moreover, there is an OSP mechanism with monitoring that is a PTAS for the same problem. Both mechanisms satisfy voluntary participation.*

## 5 Facility Location

We now show that, without monitoring, there is no OSP mechanism for the facility location problem with an approximation ratio better than  $n - 1$ . To this aim, let us first state the following simple observation.

**Observation 1** *For every  $\alpha, \beta$ , with  $\alpha < \beta$ , no  $k$ -approximate mechanism  $\mathcal{M}$  for the facility location problem, with  $k < n - 1$ , sets  $f(\mathbf{d}) \leq \alpha$ , if  $d_i = \alpha$  and  $d_j = \beta$  for every  $j \neq i$ , and  $f(\mathbf{d}') \geq \beta$ , if  $d'_i = \beta$  and  $d'_j = \alpha$  for every  $j \neq i$ , where  $f$  is the decision policy implemented by  $\mathcal{M}$ .*

This observation will be proved below in a more general statement.

Observation 1, if combined with the characterization of OSP mechanisms in the setting of voting with single-peaked preferences given in [6], immediately holds that no approximation better than  $n - 1$  is possible without monitoring and without money. Next we prove that each of these conditions alone is sufficient for the

inapproximability result. That is, we prove that no approximation better than  $n - 1$  is possible without monitoring (cf. Theorem 3) or without money (cf. Theorem 4).

**Theorem 3** *For every  $\varepsilon > 0$ , there is no  $(n - 1 - \varepsilon)$ -approximate mechanism for the facility location problem that is OSP without monitoring.*

**Proof** Let  $\mathcal{M}$  be a  $(n - 1 - \varepsilon)$ -approximate mechanism for the facility location problem that is OSP without monitoring and let  $(f, p)$  be the social choice function that this mechanism implements. Let us restrict the domain of every agent to  $D' = \{a, a + \delta, \dots, b - \delta, b\}$ , where  $\delta \leq \frac{\varepsilon}{n-2} \cdot \frac{b-a}{2}$ . Let  $\mathcal{M}'$  be the mechanism obtained by pruning  $\mathcal{M}$  according to this restriction. As stated above,  $\mathcal{M}'$  must be an OSP mechanism. Moreover, the approximation ratio of  $\mathcal{M}'$  cannot be worse than the approximation ratio of  $\mathcal{M}$ . Hence,  $\mathcal{M}'$  cannot be trivial, otherwise its approximation ratio would be unbounded.

Then, let  $i$  be the divergent agent of  $\mathcal{M}'$ . Note that, by definition of divergent agent, there must be two types  $t_i, t'_i$  of agent  $i$  such that  $t'_i = t_i + \delta$  and  $i$  takes an action in  $\mathcal{M}'$  when her type is  $t_i$  that is different from the action taken when her type is  $t'_i$ . We denote as  $c$  and  $d$  the smallest  $t_i$  and the largest  $t'_i$ , respectively, for which this occurs, i.e.,  $c$  is the smallest type in  $D'$  such that  $i$  diverges on  $c$  and  $c + \delta$ , and  $d$  is the largest type in  $D'$  such that  $i$  diverges on  $d$  and  $d - \delta$ .

Note that either  $c < \frac{b+a}{2}$  or  $d > \frac{b+a}{2}$ . Indeed, if  $c \geq \frac{b+a}{2}$ , then  $d \geq c + \delta > \frac{b+a}{2}$ . In the rest of the proof we will assume that  $c < \frac{a+b}{2}$ . The proof for the case that  $d > \frac{a+b}{2}$  simply requires to replace  $c$  with  $d$ ,  $c + \delta$  with  $d - \delta$ , and  $b$  with  $a$ , and invert the direction of the inequalities in the next claims.

The proof uses two profiles  $\mathbf{x}$  and  $\mathbf{y}$ , that are defined as follows:

- $x_i = c + \delta$ , and  $x_k = c$  for every  $k \neq i$ ;
- $y_i = c$ , and  $y_k = b$  for every  $k \neq i$ .

We begin by using OSP to relate payments and outcomes of the mechanism  $\mathcal{M}'$  on input  $\mathbf{x}$  and  $\mathbf{y}$ . Specifically, we note that if the real location of  $i$  is  $t_i = x_i = c + \delta$  then  $c_i(t_i, \mathcal{M}'(\mathbf{x})) = d(c + \delta, f(\mathbf{x})) - p_i(\mathbf{x})$ , and  $c_i(t_i, \mathcal{M}'(\mathbf{y})) = d(c + \delta, f(\mathbf{y})) - p_i(\mathbf{y})$ . Since  $i$  diverges on  $c$  and  $c + \delta$  and  $\mathcal{M}'$  is OSP, we have that  $c_i(t_i, \mathcal{M}'(\mathbf{x})) \leq c_i(t_i, \mathcal{M}'(\mathbf{y}))$ . Hence, it follows that

$$p_i(\mathbf{x}) \geq p_i(\mathbf{y}) - d(c + \delta, f(\mathbf{y})) + d(c + \delta, f(\mathbf{x})). \quad (1)$$

Suppose instead that the real location of  $i$  is  $t'_i = y_i = c$  then  $c_i(t'_i, \mathcal{M}'(\mathbf{x})) = d(c, f(\mathbf{x})) - p_i(\mathbf{x})$ , and  $c_i(t'_i, \mathcal{M}'(\mathbf{y})) = d(c, f(\mathbf{y})) - p_i(\mathbf{y})$ . As above, since  $i$  diverges on  $c$  and  $c + \delta$  and  $\mathcal{M}'$  is OSP, we have that  $c_i(t'_i, \mathcal{M}'(\mathbf{y})) \leq c_i(t'_i, \mathcal{M}'(\mathbf{x}))$ . Hence, it follows that

$$p_i(\mathbf{x}) \leq p_i(\mathbf{y}) - d(c, f(\mathbf{y})) + d(c, f(\mathbf{x})). \quad (2)$$

Therefore, in order to satisfy both (1) and (2), we need that

$$d(c + \delta, f(\mathbf{y})) - d(c, f(\mathbf{y})) \geq d(c + \delta, f(\mathbf{x})) - d(c, f(\mathbf{x})). \quad (3)$$

Using (3) above, we first show that  $f(\mathbf{x})$  must be at least  $c$  and then that  $f(\mathbf{y}) \leq c + \delta$ . Finally, we prove how this last fact contradicts the desired approximation ratio.

Let us first show that  $f(\mathbf{x}) \geq c$ . Suppose instead that  $f(\mathbf{x}) < c$ . Since  $f(\mathbf{x}) < c$ , then the r.h.s. of (3) is  $\delta$ . As for the l.h.s., we distinguish two cases. If  $f(\mathbf{y}) \leq c + \delta$ , then, since  $f(\mathbf{y}) > c$  according to Observation 1, then we have  $(c + \delta - f(\mathbf{y})) - (f(\mathbf{y}) - c) = \delta - 2(f(\mathbf{y}) - c) < \delta$ . If  $f(\mathbf{y}) > c + \delta$ , we have  $(f(\mathbf{y}) - (c + \delta)) - (f(\mathbf{y}) - c) = -\delta$ . Hence, in both cases we reach a contradiction.

We now show that  $f(\mathbf{y}) \leq c + \delta$ . Assume by contradiction that  $f(\mathbf{y}) > c + \delta$ . Since  $f(\mathbf{x}) \geq c$ , and  $f(\mathbf{x}) < c + \delta$  by Observation 1, we can rewrite (3) as

$$(f(\mathbf{y}) - (c + \delta)) - (f(\mathbf{y}) - c) \geq ((c + \delta) - f(\mathbf{x})) - (f(\mathbf{x}) - c) \Rightarrow -\delta \geq \delta - 2(f(\mathbf{x}) - c).$$

However, this is impossible since  $f(\mathbf{x}) < c + \delta$ .

Finally, we prove that, given that  $f(\mathbf{y}) \leq c + \delta$ , then the mechanism is not  $(n - 1 - \epsilon)$ -approximate. Indeed, since by Observation 1  $f(\mathbf{y}) > c$ , the total cost of mechanism  $\mathcal{M}'$  on input  $\mathbf{y}$  is

$$\begin{aligned} (f(\mathbf{y}) - c) + (n - 1)(b - f(\mathbf{y})) &= (n - 1)b - c - (n - 2)f(\mathbf{y}) \\ &\geq (n - 1)(b - c) - (n - 2)\delta \\ &\geq (n - 1)(b - c) - (n - 2)\frac{\epsilon}{n - 2} \cdot \frac{b - a}{2} \\ &> (n - 1 - \epsilon)(b - c), \end{aligned}$$

where we used that  $b - c > b - \frac{b+a}{2} = \frac{b-a}{2}$ . However, this is absurd, since  $\mathcal{M}'$  is  $(n - 1 - \epsilon)$ -approximate and the optimal mechanism on input  $\mathbf{y}$  places the facility in  $b$  and has total cost  $b - c$ .  $\square$

Next show that if we insist on mechanisms without money, then there is no OSP mechanism that can guarantee an approximation ratio better than  $n - 1$  even when the mechanism can use monitoring.

**Theorem 4** *For every  $\epsilon > 0$ , there is no  $(n - 1 - \epsilon)$ -approximate mechanism without money for the facility location problem that is OSP, even with monitoring.*

In order to prove Theorem 4, we first need to state the following lemma, that can be seen as a quantitative version of Observation 1.

**Lemma 1** *Consider a type profile  $\mathbf{d}$  such that  $d_i = x$  for some  $i$  and  $d_j = x - \alpha$  for every  $j \neq i$ . Then  $f(\mathbf{d}) \in \left[ x - \alpha \left( 1 + \frac{k-1}{n} \right), x - \alpha \left( 1 - \frac{k-1}{n-2} \right) \right]$  for every  $k$ -approximate mechanism  $\mathcal{M}$ , where  $f$  is the decision policy implemented by  $\mathcal{M}$ .*

**Proof** The optimal facility location for the given setting consists in placing the facility in position  $x - \alpha$ . The total cost in this case is  $\alpha$ .

If  $f(\mathbf{d}) < x - \alpha \left( 1 + \frac{k-1}{n} \right)$ , then the total cost is larger than  $(n - 1) \frac{(k-1)\alpha}{n} + \alpha + \frac{(k-1)\alpha}{n} = k\alpha$ , thus no  $k$ -approximate mechanism can place the



facility in  $f(\mathbf{d})$ . Similarly, if  $f(\mathbf{d}) > x - \alpha \left(1 - \frac{k-1}{n-2}\right)$ , then the total cost is  $(n-1)(f-x+\alpha) + x-f = (n-2)(f-x) + (n-1)\alpha > k\alpha$ , thus no  $k$ -approximate mechanism can place the facility in  $f(\mathbf{d})$ .  $\square$

We are now ready to prove Theorem 4.

**Proof of Theorem 4** Suppose there is an OSP mechanism  $\mathcal{M}$  that is  $(n-1-\varepsilon)$ -approximate. Clearly, the mechanism is non-trivial, otherwise its approximation ratio would be unbounded. Then, let  $i$  be the divergent agent of  $\mathcal{M}$ , and let  $x_i$  and  $y_i$  be the types in which  $i$  diverges. W.l.o.g., assume that  $x_i > y_i$ . Let  $\lambda = 2(x_i - y_i)$  and  $\alpha = \lambda \cdot \frac{n-2}{\varepsilon}$ . Let  $x_i$  be the truthful position of this agent. If  $i$  plays truthfully, then she can face the setting in which the remaining  $n-1$  agents are in position  $x_i - \alpha$ . By applying Lemma 1 with  $k = n-1-\varepsilon$  and  $x = x_i$ , we have that the distance of agent  $i$  from the facility must be at least  $x_i - x_i + \alpha \left(1 - \frac{n-2-\varepsilon}{n-2}\right) = \alpha \cdot \frac{\varepsilon}{n-2} = \lambda$ .

Suppose that instead  $i$  plays as if her real location would be  $y_i$ . It may be then the case that the remaining  $n-1$  agents are exactly in the same position. Then, any mechanism with bounded approximation must place the facility in  $y_i = x_i - \frac{\lambda}{2}$ . Recall that, with monitoring, the cost of agent  $i$  must be taken as the maximum between the distance to the facility either from the real position or from the declared position. In this case, this is given by the former distance and it is  $\frac{\lambda}{2} < \lambda$ . Thus, the best cost paid by  $i$  by not playing truthfully is lower than the worst cost that she can pay by playing truthfully. Then, the mechanism  $\mathcal{M}$  is not OSP, contradicting our hypothesis.

The bounds above are tight, since there is a  $(n-1)$ -approximate mechanism without money for the facility location problem that is OSP, even without monitoring. Consider, indeed, the dictatorship mechanism, in which only the dictator  $i$  is queried for her position. It is well-known that this mechanism is  $(n-1)$ -approximate. Moreover, it is also OSP [32]. Indeed, agent  $i$  is the only agent that is involved in a decision and it is always better for her to reveal her real position  $x_i$ : in this case the facility will be located exactly in her position and the cost of  $i$  will be 0, whereas by declaring a different position  $x \neq x_i$  the cost will be  $|x - x_i| > 0$ .

## 5.1 Optimal OSP Mechanisms for Facility Location

Interestingly, the combination of monitoring and monetary transfers gives an enormous power in this setting. Indeed, since the optimal facility location is a median among the positions declared by agents, it follows from Theorem 1 that there is an OSP mechanism with monitoring that computes the optimal facility location in polynomial time.

Recall that in this mechanism the agents receive a payment. As noted in the introduction, however, for a facility location problem we might need alternative mechanisms in which the designer does not pay agents, but the agents pay the mechanism. Note that this is more natural in settings wherein agents' payments can be easily

implemented via subscription fees or delayed access to the facility. We next present such an alternative optimal OSP mechanism.

We are going to assume that we are given some bounds on the agents' potential locations. (Note that in some of the related literature on facility location, agents can declare any location in  $\mathbb{R}$ .) To simplify the notation, we assume that  $D_i = [a, b]$  for all agents  $i$ . Consider now the following direct-revelation mechanism, that we call *interval mechanism*:

1. Query agents for their position in any order.
2. Let  $\mathbf{x}$  be the profile of the collected positions. Then fix the location  $f(\mathbf{x})$  of the facility to be the median of  $\mathbf{x}$ . In case of multiple medians, the facility is located on the leftmost median.
3. For every agent  $i = 1, \dots, n$ , set  $p_i(\mathbf{x}) = d(x_i, f(\mathbf{x})) - (b - a)$ .

It is not hard to see that the interval mechanism simply “shifts” the payments of the mechanism in Theorem 1 to make them of the right sign. Hence, the result below simply follows as a corollary of Theorem 1 by observing that, just like truthfulness, OSP is preserved when these shifts are bid-independent.

**Corollary 2** *The interval mechanism is an optimal mechanism that is OSP with monitoring.*

We here include a full proof of Corollary 2 for sake of completeness.

**Proof** We will next prove that the mechanism is OSP, and thus each agent has an incentive to declare her real position. Since the mechanism places the facility in the median of these positions, that is known to be the minimum cost allocation [37], it then turns out to be optimal as well.

In order to prove that it is OSP, recall that in a mechanism with monitoring the cost that  $i$  pays is  $c_i(x_i, \mathcal{M}(\mathbf{y})) = \max\{d(x_i, f(\mathbf{y})), d(y_i, f(\mathbf{y}))\} - p_i(\mathbf{y})$ . Consider then agent  $i$  and let  $x_i$  be her real position. If  $i$  declares the real position, then her total cost will be  $b - a$ . If  $i$  declares a different position  $x'_i$ , then there are two cases: if  $\min_{\mathbf{x}_{-i}} c_i(x_i, \mathcal{M}(\mathbf{x}'))$  is achieved in a profile  $\mathbf{x}'_{-i}$  such that  $f(\mathbf{x}') \neq x'_i$ , then

$$\begin{aligned} c_i(x_i, \mathcal{M}(\mathbf{x}')) &= \max\{d(x_i, f(\mathbf{x}')), d(x'_i, f(\mathbf{x}'))\} - p_i(\mathbf{x}') \\ &\geq d(x'_i, f(\mathbf{x}')) - p_i(\mathbf{x}') = b - a; \end{aligned}$$

otherwise (that is, if  $f(\mathbf{x}') = x'_i \neq x_i$ )

$$\begin{aligned} c_i(x_i, \mathcal{M}(\mathbf{x}')) &= \max\{d(x_i, f(\mathbf{x}')), d(x'_i, f(\mathbf{x}'))\} - p_i(\mathbf{x}') \\ &= d(x_i, x'_i) - p_i(\mathbf{x}') > b - a. \end{aligned}$$

Thus, in both cases the best cost that  $i$  can obtain by declaring a position different from the real one is not smaller than the worst cost that  $i$  can obtain by playing truthfully.  $\square$

The drawback of the interval mechanism is that the payment that this mechanism charges may be as large as the size of the interval. This opens the question of whether more frugal payment schemes exist—or in other words, how susceptible OSP with monitoring is to payment shifts that are not bid independent.

Next section will show that it is indeed possible to slightly optimize the interval mechanism in order to be less expensive for the agents. We further prove (in Sect. 5.3) that our optimization is optimal as long as we focus on direct-revelation mechanisms. However, even this optimized version still requires that  $O(n)$  agents will pay an amount that is about  $b - a$ . We will finally show (in Sect. 5.4) that this is unavoidable for at least one type profiles, even if one considers mechanisms that are not direct-revelation.

## 5.2 The Optimized Interval Mechanism

Consider the following optimized version of the interval mechanism, that we call *Optimized Interval Mechanism (OIM)*:

1. Query agents for their position in any *sequential* order.
2. Let  $\mathbf{x}$  be the profile of the collected positions. Then fix the location  $f(\mathbf{x})$  of the facility to be the median of  $\mathbf{x}$ . In case of multiple medians, the facility is located on the leftmost median.
3. For every agent  $i \in [n]$ , let  $K_i(k_i, \text{resp.})$  be the set (number, resp.) of agents queried before  $i$ . Let  $\mathbf{s} = (s_1, \dots, s_{k_i})$  be the profile containing the locations declared by these agents in non-decreasing order. Let  $\ell = \left\lceil \frac{n}{2} \right\rceil + k_i - n + 1$  and  $r = \left\lfloor \frac{n}{2} \right\rfloor - 1$ . If  $\ell > 1$ ,  $r < k_i$  and  $s_{\ell-1} = s_{r+1}$ , then set  $p_i(\mathbf{x}) = 0$ . Otherwise, we define  $L_i$  and  $R_i$  as follows:

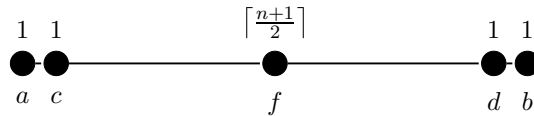
$$L_i = \begin{cases} s_\ell, & \text{if } \ell \geq 1; \\ a, & \text{otherwise.} \end{cases} \quad R_i = \begin{cases} s_r, & \text{if } r \leq k_i; \\ b, & \text{otherwise.} \end{cases}$$

Let also define  $A_i$  and  $B_i$  as follows:

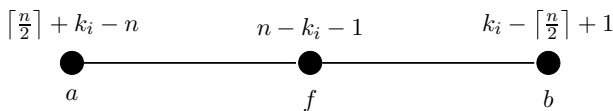
$$A_i = \begin{cases} L_i, & \text{if } L_i = R_i; \\ 2L_i - b, & \text{if } R_i > L_i > \frac{a+b}{2}; \\ a, & \text{otherwise.} \end{cases} \quad B_i = \begin{cases} R_i, & \text{if } L_i = R_i; \\ 2R_i - a, & \text{if } L_i < R_i < \frac{a+b}{2}; \\ b, & \text{otherwise.} \end{cases}$$

Finally, let  $m_i = \max\{R_i - A_i, B_i - L_i\}$ . If  $x_i \in [A_i, B_i]$ , then set  $p_i(\mathbf{x}) = d(x_i, f(\mathbf{x})) - m_i$ . If  $x_i < A_i$ , then set  $p_i(\mathbf{x}) = d(x_i, f(\mathbf{x})) - m_i - d(x_i, A_i)$ . If  $x_i > B_i$ , then set  $p_i(\mathbf{x}) = d(x_i, f(\mathbf{x})) - m_i - d(x_i, B_i)$ .

The idea behind OIM is to exploit the information given by the extensive-form implementation of the mechanism to reduce the charge to the bidders, i.e., use the value of  $k_i$  to reduce the payment to bidder  $i$ . In fact, when all bidders bid simultaneously then  $k_i = 0$  for all  $i$  and OIM is simply the interval mechanism.



**Fig. 1** When  $i$  is queried, there are at least  $\lceil \frac{n+1}{2} \rceil$  agents that have declared the same position  $f$ . In this case the facility will be placed in  $f$  regardless of the position declared by  $i$  and other remaining agents. OIM will then pay 0 to  $i$ , whereas a larger payment would be charged by the interval mechanism



**Fig. 2** Let  $k_i \geq \lceil \frac{n}{2} \rceil$ . It is not hard to check that in this case, if  $i$  declares  $f$ , then the facility will be placed in  $f$  regardless of the position declared by the remaining  $n - k_i - 1$  agents. OIM will then pay  $|x_i - f|$  to  $i$ , whereas a larger payment would be charged by the interval mechanism

The way in which this optimization upon  $k_i$  is done can arguably appear a bit complex but is not too hard to explain. First, when  $s_{\ell-1} = s_{r+1}$  (this occurs whenever there are at least  $\lceil \frac{n+1}{2} \rceil$  among the ones queried before  $i$  that declared exactly the same position—see, e.g., Fig. 1), then the facility will be placed in  $s_{\ell-1}$  regardless of the location declared by  $i$  and by every other agent  $j$  queried after  $i$ . Hence, these agents will not have any incentive in declaring a position that is different from their real location even without payments.

As for the second and most important optimization step, we consider profiles  $\mathbf{x}$  for which there are agents very far away from the facility. Indeed, as we will hint in Lemma 2, the facility is very likely to be included in the interval  $[L_i, R_i]$ . Thus, if an agent  $i$  in  $\mathbf{x}$  is very far away from this interval, one can slightly lower the payment assigned to her and still have an OSP mechanism. More details on the effectiveness of this optimization can be found in Lemma 9.

We highlight that this last optimization is particularly relevant when there is a location  $f$  such that when  $i$  declares  $f$ , then the facility will securely be located in  $f$  even if  $s_{\ell-1} \neq s_{r+1}$  (this case corresponds to  $L_i = R_i = f$ —see, e.g., Fig. 2). In this case, it is possible to reduce the cost of agent  $i$  in profile  $\mathbf{x}$  from  $b - a$  to  $|x_i - f|$ .

Nevertheless, we note that the mechanism still has very large costs, namely  $b - a$ , for at least  $\lceil \frac{n}{2} \rceil - 1$  agents. Indeed, for these agents, it turns out that  $L_i = A_i = a$  and  $R_i = B_i = b$ , and thus  $m_i = b - a$ . We will show in the next section that this is inevitable with a direct-revelation mechanism.

In order to prove that OIM is an optimal OSP mechanism, we say that a profile  $\mathbf{x}'$  is  $i$ -compatible if  $\mathbf{x}' = (\mathbf{x}_{K_i}, \mathbf{x}'_{-K_i})$ , i.e.,  $x'_j = x_j$  for all the agents  $j \in K_i$ , where  $K_i$

is the set of agents that in OIM revealed her type before  $i$ . We then start by making some observations on  $L_i$  and  $R_i$ . Specifically, in Lemma 2 we show that these quantities represent a lower and an upper bound for the position of the facility as long as the position declared by  $i$  is not below  $L_i$  or above  $R_i$ . Lemma 3 instead proves that, when this occurs, i.e. the position declared by  $i$  is not in the interval  $[L_i, R_i]$ , then it is the declared position to be a lower bound (if below  $L_i$ ) or an upper bound (if above  $R_i$ ) to the facility location.

**Lemma 2** *For every  $i$ , and every  $t > L_i$ , it holds that  $f(\mathbf{x}') \geq L_i$  for every  $i$ -compatible profile  $\mathbf{x}'$  with  $x'_i = t$ ; moreover, there is one such profile for which  $f(\mathbf{x}') = L_i$ . Similarly, for every  $t < R_i$ ,  $f(\mathbf{x}') \leq R_i$  for every  $i$ -compatible profile  $\mathbf{x}'$  with  $x'_i = t$ , and there is one such profile for which  $f(\mathbf{x}') = R_i$ .*

**Proof** Let  $t > L_i$  and consider the profile  $\mathbf{x}'$  such that

$$x'_j = \begin{cases} x_j, & \text{if } j \in K_i; \\ t, & \text{if } j = i; \\ a, & \text{otherwise.} \end{cases}$$

It is easy to see that  $f(\mathbf{x}') = L_i$ . Indeed, if  $\ell < 1$ , then in  $\mathbf{x}'$  there are  $n - k_i - 1 = \left\lfloor \frac{n}{2} \right\rfloor - \ell > \left\lfloor \frac{n}{2} \right\rfloor - 1$  agents whose location is  $a$ . Hence, the leftmost median of  $\mathbf{x}'$  must be  $a = L_i$ . If  $\ell \geq 1$ , then  $L_i$  is the  $\ell$ -th smallest location among agents that are processed before  $i$ , and there are in  $\mathbf{x}'$  exactly  $n - k_i - 1$  agents whose location is surely not larger than  $L_i$ . Then  $L_i$  is the  $\ell + n - k_i - 1 = \left\lfloor \frac{n}{2} \right\rfloor$ -th smallest location in  $\mathbf{x}'$ , i.e. the (leftmost) median.

On the other hand, it is immediate to see that there is no declaration by agents  $j \notin K_i$ , with  $j \neq i$ , that can make the facility go to the left of  $L_i$ .

Let now  $t < R_i$  and consider the profile  $\mathbf{x}'$  such that

$$x'_j = \begin{cases} x_j, & \text{if } j \in K_i; \\ t, & \text{if } j = i; \\ b, & \text{otherwise.} \end{cases}$$

It is easy to see that  $f(\mathbf{x}') = R_i$ . Indeed, if  $r > k_i$ , then in  $\mathbf{x}'$  there are  $n - k_i - 1 > \left\lfloor \frac{n}{2} \right\rfloor$  agents whose location is  $b$ . Hence, the median of  $\mathbf{x}'$  is  $b = R_i$ . If  $r \leq k_i$ , then  $R_i$  is the  $r$ -th smallest location among agents that are processed before  $i$ , and there is in  $\mathbf{x}'$  exactly one agent whose location is smaller than  $R_i$ . Then  $R_i$  is the  $r + 1 = \left\lfloor \frac{n}{2} \right\rfloor$ -th smallest location in  $\mathbf{x}'$ , i.e. the (leftmost) median.

Moreover, as above, it is immediate to see that there is no declaration by agents  $j \notin K_i$ , with  $j \neq i$ , that can make the facility go to the right of  $R_i$ .  $\square$

**Lemma 3** For every  $i$ , and every  $t \leq L_i$ ,  $f(\mathbf{x}') \geq t$  for every  $i$ -compatible profile  $\mathbf{x}'$  with  $x'_i = t$ . Similarly, for every  $t \geq R_i$ ,  $f(\mathbf{x}') \leq t$  for every  $i$ -compatible profile  $\mathbf{x}'$  with  $x'_i = t$ .

**Proof** Let  $t \leq L_i$  and consider the profile  $\mathbf{x}'$  such that

$$x'_j = \begin{cases} x_j, & \text{if } j \in K_i; \\ t, & \text{if } j = i; \\ a, & \text{otherwise.} \end{cases}$$

If  $\ell \geq 1$ , then  $L_i$  is the  $\ell$ -th smallest location among agents that are processed before  $i$ , and there are in  $\mathbf{x}'$  exactly  $n - k_i$  agents whose location is surely not larger than  $L_i$ . Then  $L_i$  is the  $\ell + n - k_i = \left\lceil \frac{n}{2} \right\rceil + 1$ -th smallest location in  $\mathbf{x}'$ . Thus, the leftmost median of  $\mathbf{x}'$  will be  $s_{\ell-1}$  if  $t \leq s_{\ell-1}$  and  $t$  otherwise. When  $\ell < 1$ , then  $L_i = a$  and clearly  $f(\mathbf{x}') \geq a$ .

If  $t \geq R_i$ , let us consider the profile  $\mathbf{x}'$  such that

$$x'_j = \begin{cases} x_j, & \text{if } j \in K_i; \\ t, & \text{if } j = i; \\ b, & \text{otherwise.} \end{cases}$$

If  $r \leq k_i$ , then  $R_i$  is the  $r$ -th smallest location among agents that are processed before  $i$ , and there is in  $\mathbf{x}'$  no agent whose location is surely smaller than  $R_i$ . Then  $R_i$  is the  $r = \left\lceil \frac{n}{2} \right\rceil - 1$ -th smallest location in  $\mathbf{x}'$ . Thus, the leftmost median of  $\mathbf{x}'$  will be  $s_{r+1}$  if  $t \geq s_{r+1}$  and  $t$  otherwise. When  $r > k_i$ , then  $R_i = b$  and clearly  $f(\mathbf{x}') \leq b$ .  $\square$

We now show that OIM is optimal and OSP with monitoring.

**Theorem 5** OIM is an optimal mechanism that is OSP with monitoring.

**Proof** We will next prove that the mechanism is OSP with monitoring, and thus for each agent it is obviously dominant to declare her real position. Since OIM places the facility on the median of these positions, that is known to be the minimum cost allocation [37], it then turns out to be optimal as well.

Consider then agent  $i$  and let  $x_i$  be her real position. If  $s_{\ell-1} = s_{r+1} = \lambda$ , then the facility will be located in  $\lambda$  and  $i$  receives a zero payment, regardless of her declaration and the declarations of the agents not in  $K_i$ .

Suppose now that  $s_{\ell-1} \neq s_{r+1}$  and the real position of  $i$  is  $x_i \in [A_i, B_i]$ . If  $i$  declares her real position, then her total cost will be at most  $m_i$ . If  $i$  declares a different position  $x'_i$ , then for every  $i$ -compatible profile  $\mathbf{x}'$

$$\begin{aligned} c_i(x_i, \mathcal{M}(\mathbf{x}')) &= \max\{d(x_i, f(\mathbf{x}')), d(x'_i, f(\mathbf{x}'))\} - p_i(\mathbf{x}') \\ &\geq d(x'_i, f(\mathbf{x}')) - p_i(\mathbf{x}') = c_i(x'_i, \mathcal{M}(\mathbf{x}')) \geq m_i. \end{aligned}$$

Suppose now that  $s_{\ell-1} \neq s_{r+1}$  and the real position of  $i$  is  $x_i = A_i - c$  or  $x_i = B_i + c$  with  $c > 0$ . W.l.o.g. we will assume  $x_i = A_i - c$ . If  $i$  declares the real position, then her total cost will be at most  $m_i + c$ . If  $i$  declares a position  $x'_i = A_i - c'$  or  $x'_i = B_i + c'$  with  $c' > c$ , then for every  $i$ -compatible profile  $\mathbf{x}'$  we have that

$$\begin{aligned} c_i(x_i, \mathcal{M}(\mathbf{x}')) &= \max\{d(x_i, f(\mathbf{x}')), d(x'_i, f(\mathbf{x}'))\} - p_i(\mathbf{x}') \geq d(x'_i, f(\mathbf{x}')) - p_i(\mathbf{x}') \\ &\geq m_i + c' > m_i + c. \end{aligned}$$

If  $i$  declares a position  $x'_i = A_i - c'$  for  $0 < c' < c$ , then for every  $i$ -compatible profile  $\mathbf{x}'$  we have that

$$\begin{aligned} c_i(x_i, \mathcal{M}(\mathbf{x}')) &= \max\{d(x_i, f(\mathbf{x}')), d(x'_i, f(\mathbf{x}'))\} - p_i(\mathbf{x}') = d(x_i, f(\mathbf{x}')) - p_i(\mathbf{x}') \\ &= d(x_i, x'_i) + d(x'_i, f(\mathbf{x}')) - p_i(\mathbf{x}') = (c - c') + m_i + c' = m_i + c, \end{aligned}$$

where we used that, according to Lemma 3,  $f(\mathbf{x}') \geq x'_i$  and thus  $d(x_i, f(\mathbf{x}')) = d(x_i, x'_i) + d(x'_i, f(\mathbf{x}'))$ .

If  $i$  instead declares a position  $x'_i \in [A_i, B_i + c]$ , then for every  $i$ -compatible profile  $\mathbf{x}'$  such that  $f(\mathbf{x}') \geq x'_i$  we have that

$$\begin{aligned} c_i(x_i, \mathcal{M}(\mathbf{x}')) &= \max\{d(x_i, f(\mathbf{x}')), d(x'_i, f(\mathbf{x}'))\} - p_i(\mathbf{x}') = d(x_i, f(\mathbf{x}')) - p_i(\mathbf{x}') \\ &= d(x_i, x'_i) + d(x'_i, f(\mathbf{x}')) - p_i(\mathbf{x}') \geq m_i + c. \end{aligned} \quad (4)$$

For every  $i$ -compatible profile  $\mathbf{x}'$  such that  $f(\mathbf{x}') < x'_i$ , we have instead that

$$\begin{aligned} c_i(x_i, \mathcal{M}(\mathbf{x}')) &= \max\{d(x_i, f(\mathbf{x}')), d(x'_i, f(\mathbf{x}'))\} - p_i(\mathbf{x}') = d(x_i, f(\mathbf{x}')) - p_i(\mathbf{x}') \\ &= d(x_i, f(\mathbf{x}')) - d(x'_i, f(\mathbf{x}')) + m_i + \max\{0, x'_i - B_i\} \\ &= f(\mathbf{x}') - x_i - x'_i + f(\mathbf{x}') + m_i + \max\{0, x'_i - B_i\} \\ &\geq (L_i - x_i) - (x'_i - L_i) + m_i + \max\{0, x'_i - B_i\} \\ &= (L_i - A_i) + (A_i - x_i) - (x'_i - B_i) - (B_i - L_i) + m_i \\ &\quad + \max\{0, x'_i - B_i\} \\ &\geq m_i + c, \end{aligned} \quad (5)$$

where we used that, according to Lemma 2,  $f(\mathbf{x}') \geq L_i$ , and that  $A_i > a$ , and therefore

$$d(x'_i, f(\mathbf{x}')) \leq B_i + c - L_i = L_i - A_i + c = L_i - x_i \leq d(x_i, f(\mathbf{x}')).$$

□

### 5.3 Payments in Direct-Revelation Mechanisms are at least as in OIM

Next we will prove that no direct-revelation mechanism in which the mechanism does not pay the agents can be optimal for the facility location problem and OSP with monitoring with smaller payments than OIM.

Specifically, given two optimal mechanisms  $\mathcal{M}$  and  $\mathcal{M}'$  implementing the same decision rule, with  $p$  and  $p'$  being the corresponding implemented payment policies, and an agent  $i$  we say that  $\mathcal{M}$  is less *demanding* to  $i$  than  $\mathcal{M}'$  if the sum over all type profiles of the payments that  $\mathcal{M}$  assigns to  $i$  is less than the same sum computed on the payments assigned by  $\mathcal{M}'$ , i.e.,  $\sum_{\mathbf{x}} p_i(\mathbf{x}) < \sum_{\mathbf{x}} p'_i(\mathbf{x})$ .

Next theorem then formalize that OIM is the less demanding mechanism within the class of direct-revelation mechanism. Note that henceforth we only consider mechanisms that adopt payments different from OIM, i.e., there is at least one profile  $\mathbf{x}$  and one agent  $i$  such that the payment assigned at  $i$  on profile  $\mathbf{x}$  is different from the corresponding payment assigned by OIM.

**Theorem 6** *For every optimal OSP direct-revelation mechanism  $\mathcal{M}$  for the facility location problem, if the payment policy satisfies that  $p_i(\mathbf{x}) \leq 0$  for every agent  $i$  and every profile  $\mathbf{x}$  (i.e., the mechanism does not pay the agents), then there is an agent  $i$  such that OIM is less demanding to  $i$  than  $\mathcal{M}$ .*

**Proof** Fix a player  $i$  and recall that  $K_i$  is the set of agents whose location is known to  $i$  when she is queried. Let  $\mathcal{M}$  be an optimal OSP direct-revelation mechanism for facility location with payment policy  $p$ , and, assume without loss of generality that it is the mechanism that demands less to  $i$ , i.e.,  $\sum_{\mathbf{x}} p_i(\mathbf{x}) = \arg \min_{(f,p)} \text{OSP} \sum_{\mathbf{x}} p_i(\mathbf{x})$ .

Assume, by contradiction, that  $\mathcal{M}$  is less demanding than OIM. Clearly, it cannot set lower non-positive payments than OIM when  $s_{\ell-1} = s_{r+1}$ . Thus, we can safely consider that  $s_{\ell-1} \neq s_{r+1}$ . Next lemmata show some conditions that payments must satisfy in order for  $\mathcal{M}$  to be OSP and to minimize the sum of payments assigned to  $i$ . Specifically, Lemmas 4 and 5 focus on profiles  $\mathbf{x}$  such that  $x_i \in [A_i, B_i]$  and  $L_i \neq R_i$ , Lemma 6 consider profiles  $\mathbf{x}$  such that  $x_i = L_i = R_i$ , whereas Lemma 7 focuses on the remaining profiles.

Suppose first that  $L_i \neq R_i$ . Then let  $\mu_i$  be the minimum cost that  $i$  pays in a profile  $\mathbf{x}$  such that  $f(\mathbf{x}) \neq x_i$  (such a profile surely exists since, by optimality of  $\mathcal{M}$ ,  $i$  is not a dictator), i.e.,  $\mu_i = \min_{\mathbf{x}: f(\mathbf{x}) \neq x_i} c_i(x_i, \mathcal{M}(\mathbf{x}))$ . We begin by proving this useful claim. □

**Claim 1** *Let  $\mathbf{x}$  be a profile such that  $f(\mathbf{x}) \neq x_i$  and  $c_i(x_i, \mathcal{M}(\mathbf{x})) = \mu_i$ . If  $x_i < f(\mathbf{x})$ , then for every  $y_i \in (x_i, \min\{2f(\mathbf{x}) - x_i, b\}]$ , it turns out that  $c_i(y_i, \mathcal{M}(\mathbf{y})) = \mu_i$  if  $f(\mathbf{y}) \neq y_i$ , and  $c_i(y_i, \mathcal{M}(\mathbf{y})) \leq \mu_i$  otherwise.*

*Similarly, if  $x_i > f(\mathbf{x})$ , then for every  $y_i \in [\max\{a, 2f(\mathbf{x}) - x_i\}, x_i)$ , it turns out that  $c_i(y_i, \mathcal{M}(\mathbf{y})) = \mu_i$  if  $f(\mathbf{y}) \neq y_i$ , and  $c_i(y_i, \mathcal{M}(\mathbf{y})) \leq \mu_i$  otherwise.*

**Proof** Since  $\mathcal{M}$  is a direct-revelation mechanism, then  $i$  diverges on  $y_i$  and  $x_i$ . Then, since  $\mathcal{M}$  is OSP, it must be the case that

$$\begin{aligned} c_i(y_i, \mathcal{M}(\mathbf{y})) &\leq c_i(y_i, \mathcal{M}(\mathbf{x})) = \max\{d(y_i, f(\mathbf{x})), d(x_i, f(\mathbf{x}))\} - p_i(\mathbf{x}) \\ &= d(x_i, f(\mathbf{x})) - p_i(\mathbf{x}) = c_i(x_i, \mathcal{M}(\mathbf{x})) = \mu_i, \end{aligned}$$

where we used that  $d(y_i, f(\mathbf{x})) \leq d(x_i, f(\mathbf{x}))$  by definition of  $y_i$ .



However, if  $f(\mathbf{y}) \neq y_i$ , then, by definition of  $\mu_i$ , it must be the case that  $c_i(y_i, \mathcal{M}(\mathbf{y})) \geq \mu_i$ , that leaves  $c_i(y_i, \mathcal{M}(\mathbf{y})) = \mu_i$  as the only possible option.  $\square$

**Lemma 4** *If  $L_i \neq R_i$ , then for every  $y_i \in [A_i, B_i]$ ,  $c_i(y_i, \mathcal{M}(\mathbf{y})) = \mu_i$  if  $f(\mathbf{y}) \neq y_i$ , and  $c_i(y_i, \mathcal{M}(\mathbf{y})) \leq \mu_i$  otherwise.*

**Proof** Consider the following procedure:

1. Let  $w = 1$ ,  $\Delta^0 = \emptyset$ , and  $\mathbf{y}^0$  be a profile achieving cost  $\mu_i$ .
2. Let  $t^w = 2f(\mathbf{y}^{w-1}) - y_i^{w-1}$ .
3. If  $t^w \geq \frac{L_i + R_i}{2}$ , consider the profile  $\mathbf{y}^w$  such that

$$y_j^w = \begin{cases} x_j, & \text{if } j \in K_i; \\ t^w, & \text{if } j = i; \\ a, & \text{otherwise.} \end{cases}$$

and let  $\Delta^w = [\max\{a, 2f(\mathbf{y}^w) - t^w\}, t^w]$ . Otherwise consider  $\mathbf{y}^w$  such that

$$y_j^w = \begin{cases} x_j, & \text{if } j \in K_i; \\ t^w, & \text{if } j = i; \\ b, & \text{otherwise.} \end{cases}$$

and let  $\Delta^w = [t^w, \min\{2f(\mathbf{y}^w) - t^w, b\}]$ .

4. If  $[A_i, B_i] \not\subseteq \Delta^w$ , set  $w = w + 1$  and repeat from step 5.3.

Let us first prove, by induction, that for every  $w \geq 0$ , it holds that  $c_i(y_i^w, \mathcal{M}(\mathbf{y}^w)) = \mu_i$ . This is clearly true for  $w = 0$ . Suppose now that  $c_i(y_i^{w-1}, \mathcal{M}(\mathbf{y}^{w-1})) = \mu_i$ . If  $t^w \geq \frac{L_i + R_i}{2}$ , then, according to Lemma 2, we have that  $f(\mathbf{y}^w) = L_i \neq y_i^w$ . Similarly, if  $t^w < \frac{L_i + R_i}{2}$ , then, according to Lemma 2, we have that  $f(\mathbf{y}^w) = R_i \neq y_i^w$ . Then, by Claim 1 applied with  $\mathbf{x} = \mathbf{y}^{w-1}$ , it holds that  $c_i(y_i^w, \mathcal{M}(\mathbf{y}^w)) = \mu_i$ . In fact, Claim 1 actually proves that for every  $w \geq 0$ , and every  $y_i \in \Delta^w$ , it holds that  $c_i(y_i, \mathcal{M}(\mathbf{y})) = \mu_i$  if  $f(\mathbf{y}) \neq y_i$ , and  $c_i(y_i, \mathcal{M}(\mathbf{y})) \leq \mu_i$  otherwise. Hence, we are only left to prove that there is a  $w$  such that  $\Delta^w \supseteq [A_i, B_i]$ .

To this aim, we next we prove that for every  $w \geq 1$ , the size of the range  $\Delta^w$  is larger than the size of the range  $\Delta^{w-1}$ . This is clearly true for  $w = 1$  since  $|\Delta^0| = 0$  and  $f(\mathbf{y}^1) \neq t^1$ , from which we achieve that  $|\Delta^1| \geq 2|f(\mathbf{y}^1) - t^1| > 0$ . Consider, instead,  $w > 1$ . Suppose that  $\Delta^{w-1} = [\max\{a, 2f(\mathbf{y}^{w-1}) - t^{w-1}\}, t^{w-1}]$ , from which we have that  $|\Delta^{w-1}| \geq 2t^{w-1} - 2f(\mathbf{y}^{w-1})$ . Note that this only occurs if  $t^{w-1} \geq \frac{L_i + R_i}{2} = L_i + \frac{R_i - L_i}{2}$  and thus  $f(\mathbf{y}^{w-1}) = L_i$ , from which it follows that  $t^w = 2f(\mathbf{y}^{w-1}) - t^{w-1} \leq L_i - \frac{R_i - L_i}{2} < \frac{L_i + R_i}{2}$ . Therefore  $\Delta^w = [t^w, \min\{2f(\mathbf{y}^w) - t^w, b\}]$ . If  $t_w \geq 2R_i - b$ , then  $\min\{2f(\mathbf{y}^w) - t^w, b\} \neq b$ , and thus  $|\Delta_w| = 2f(\mathbf{y}^w) - 2t^w = 2R_i - 2t_w$ , otherwise  $|\Delta_w| = b - t^w \geq 2R_i - 2t_w$ . Hence, in both cases we achieve that

$$|\Delta^w| \geq 2R_i - 2f(\mathbf{y}^{w-1}) + 2t^{w-1} - 2f(\mathbf{y}^{w-1}) = 2t^{w-1} - 2f(\mathbf{y}^{w-1}) \\ + 2|R_i - L_i| > |\Delta^{w-1}|.$$

The case for  $\Delta^{w-1} = [t^{w-1}, \min\{b, 2f(\mathbf{y}^{w-1}) - t^{w-1}\}]$  can be similarly proved.

The lemma then follows since the above procedure eventually considers  $\Delta \supseteq [A_i, B_i]$ .

**Lemma 5** *If  $L_i \neq R_i$ , then for every  $y_i \in [A_i, B_i]$  and every  $i$ -compatible profile  $\mathbf{y}$  such that  $f(\mathbf{y}) = y_i$ , it holds that  $c_i(y_i, \mathcal{M}(\mathbf{y})) = \mu_i$ .*

**Proof** Suppose that there is an  $i$ -compatible profile  $\mathbf{y}$  with  $f(\mathbf{y}) = y_i$  and  $c_i(y_i, \mathcal{M}(\mathbf{y})) < \mu_i$ . Consider  $\delta < \min\{\mu_i - c_i(y_i, \mathcal{M}(\mathbf{y})), \max\{y_i - A_i, B_i - y_i\}\}$ . According to this choice of  $\delta$ , it must exist  $t \in [A_i, B_i]$  such that  $d(y_i, t) = \delta$ . Moreover, since  $L_i \neq R_i$  by hypothesis, either  $t \neq L_i$  or  $t \neq R_i$ . Then, according to Lemma 2, there is a profile  $\mathbf{y}'$  such that  $y'_i = t$  and  $f(\mathbf{y}') \neq y'_i$ . Thus, by Lemma 4, it holds that  $c_i(y'_i, \mathcal{M}(\mathbf{y}')) = \mu_i$ .

However, since  $\mathcal{M}$  is a direct-revelation mechanism, then  $i$  diverges on  $y_i$  and  $y'_i$ . Then, since  $\mathcal{M}$  is OSP, it must be the case that

$$\mu_i = c_i(y'_i, \mathcal{M}(\mathbf{y}')) \leq c_i(y'_i, \mathcal{M}(\mathbf{y})) = \max\{d(y'_i, f(\mathbf{y})), d(y_i, f(\mathbf{y}))\} - p_i(\mathbf{y}) \\ = d(y'_i, y_i) + d(y_i, f(\mathbf{y})) - p_i(\mathbf{y}) = \delta + c_i(y_i, \mathcal{M}(\mathbf{y})) < \mu_i,$$

that is absurd.  $\square$

Consider now the case that  $L_i = R_i$ . In this case we let  $\mu_i = \min_{\mathbf{x}: \mathbf{x}_i = L_i} c_i(x_i, \mathcal{M}(\mathbf{x}))$ . Next we prove a lower bound for  $\mu_i$ .

**Lemma 6** *Let  $\mathbf{y}'$  be the  $i$ -compatible profile such that  $f(\mathbf{y}') \neq y'_i$  of minimum cost, i.e.,  $\mathbf{y}' = \arg \min_{\mathbf{x}: f(\mathbf{x}) \neq \mathbf{x}_i} c_i(x_i, \mathcal{M}(\mathbf{x}))$ . If  $L_i = R_i$ , then for every  $i$ -compatible profile  $\mathbf{y}$  such that  $y_i = L_i$ , it holds that  $c_i(y_i, \mathcal{M}(\mathbf{y})) \geq c_i(y'_i, \mathcal{M}(\mathbf{y}')) - d(y_i, y'_i)$ .*

**Proof** Suppose that there is an  $i$ -compatible profile  $\mathbf{y}$  with  $y_i = L_i$  and  $c_i(y_i, \mathcal{M}(\mathbf{y})) < c_i(y'_i, \mathcal{M}(\mathbf{y}')) - d(y_i, y'_i)$ . Observe that, since  $L_i = R_i$ , it must be the case that  $f(\mathbf{y}) = y_i$ .

Since  $\mathcal{M}$  is a direct-revelation mechanism, then  $i$  diverges on  $y_i$  and  $y'_i$ . Then, since  $\mathcal{M}$  is OSP, it must be the case that

$$c_i(y'_i, \mathcal{M}(\mathbf{y}')) \leq c_i(y'_i, \mathcal{M}(\mathbf{y})) = \max\{d(y'_i, f(\mathbf{y})), d(y_i, f(\mathbf{y}))\} - p_i(\mathbf{y}) \\ = d(y'_i, y_i) + d(y_i, f(\mathbf{y})) - p_i(\mathbf{y}) < c_i(y'_i, \mathcal{M}(\mathbf{y}')),$$

that is absurd.  $\square$

**Lemma 7** *For every  $c > 0$ , if  $y_i = A_i - c$ , then for every  $i$ -compatible profile  $\mathbf{y}$   $c_i(y_i, \mathcal{M}(\mathbf{y})) \leq \mu_i + c$ .*

If  $y_i = B_i + c$ , then for every  $i$ -compatible profile  $\mathbf{y}$   $c_i(y_i, \mathcal{M}(\mathbf{y})) \leq \mu_i + c$ .

**Proof** Consider  $y_i < A_i$ , and let  $\mathbf{x}$  be an  $i$ -compatible profile with  $x_i = A_i$  with  $c_i(x_i, \mathcal{M}(\mathbf{x})) = \mu_i$  (it exists by Lemma 5 if  $L_i \neq R_i$ , and by definition of  $\mu_i$ , otherwise). By definition of  $A_i$ , we have that  $x_i = A_i \leq L_i$ . Then, by Lemma 3, we have that  $f(\mathbf{x}) \geq x_i$ , and thus  $d(y_i, f(\mathbf{x})) = d(y_i, x_i) + d(x_i, f(\mathbf{x})) = c + d(x_i, f(\mathbf{x})) > d(x_i, f(\mathbf{x}))$ .

However, since  $\mathcal{M}$  is a direct-revelation mechanism, then  $i$  diverges on  $y_i$  and  $x_i$ . Then, since  $\mathcal{M}$  is OSP, it must be the case that

$$\begin{aligned} c(y_i, f(\mathbf{y})) &\leq c_i(y_i, \mathcal{M}(\mathbf{x})) = \max\{d(y_i, f(\mathbf{x})), d(x_i, f(\mathbf{x}))\} - p_i(\mathbf{x}) \\ &= c + d(x_i, f(\mathbf{x})) - p_i(\mathbf{x}) = c + \mu_i. \end{aligned}$$

The case for  $y_i = B_i + c$  is similar.  $\square$

These lemmata fix the payments for  $i$ -compatible profiles  $\mathbf{y}$  such that  $y_i \in [A_i, B_i]$  when  $L_i \neq R_i$ . As for the remaining cases, next we show how to choose the payments that enable the mechanism  $\mathcal{M}$  to be optimal, OSP with monitoring, and to minimize the sum of payments assigned to  $i$ . In particular, Lemma 8 focuses on profiles  $\mathbf{y}$  such that  $y_i = L_i = R_i$ . Lemma 9 focuses instead on profiles  $\mathbf{y}$  with  $y_i = A_i - c$  or  $y_i = B_i + c$ .

**Lemma 8** Let  $\mathbf{y}'$  be the  $i$ -compatible profile of minimum cost among the ones for which  $f(\mathbf{y}') \neq y'_i$ , i.e.,  $\mathbf{y}' = \arg \min_{\mathbf{x}: f(\mathbf{x}) \neq x_i} c_i(x_i, \mathcal{M}(\mathbf{x}))$ . If there is an  $i$ -compatible profile  $\mathbf{y}$  with  $y_i = L_i = R_i$   $c_i(y_i, \mathcal{M}(\mathbf{y})) > c_i(y'_i, \mathcal{M}(\mathbf{y}')) - d(y_i, y'_i)$ , then there is another direct-revelation optimal OSP mechanism  $\mathcal{M}'$  that is less demanding to  $i$  than  $\mathcal{M}$ .

**Proof** Consider  $\mathcal{M}' = (f, \mathbf{p}')$  as  $\mathcal{M}$  except that it sets payments such that  $c'_i(y_i, \mathcal{M}'(\mathbf{y})) = d(y_i, f(\mathbf{y})) - p'_i(\mathbf{y}) = c_i(y'_i, \mathcal{M}(\mathbf{y}')) - d(y_i, y'_i)$ , and for every  $\mathbf{x}'$  such that  $x'_i \neq y_i$ , it sets payments such that  $c'_i(x'_i, \mathcal{M}'(\mathbf{x}')) - \min_{\mathbf{x}: x_i = y_i} c_i(x_i, \mathcal{M}(\mathbf{x})) = c_i(x'_i, \mathcal{M}(\mathbf{x}')) - \min_{\mathbf{x}: x_i = y_i} c_i(x_i, \mathcal{M}(\mathbf{x}))$ . Clearly,  $\mathcal{M}'$  is a direct-revelation mechanism. Moreover, since it places the facility in the median location, it is optimal if it is OSP. Finally,  $\mathcal{M}'$  reduces the payment assigned to  $i$  at least in the profile  $\mathbf{y}$ .

Hence, it is only left to show that  $\mathcal{M}'$  is OSP. Clearly, the OSP condition still holds between two profiles in which the location of  $i$  is different from  $y_i$ , and when the real location of  $i$  is exactly  $y_i$ . Next we show, that if the real location of  $i$  is  $x'_i \neq y_i$ , then it is not convenient for  $i$  to declare  $y_i$ . That is, we prove that  $c'_i(x'_i, \mathcal{M}'(\mathbf{x}')) \leq c'_i(y_i, \mathcal{M}'(\mathbf{y}))$  for every  $\mathbf{x}'$ .

Since  $L_i = R_i$ , it must be the case that  $f(\mathbf{y}) = y_i = A_i = B_i$ , and then either  $x'_i = A_i - c$  or  $x'_i = B_i + c$ , with  $c > 0$ . According to Lemma 7, we have that  $c'_i(x'_i, \mathcal{M}'(\mathbf{x}')) - c'_i(y_i, \mathcal{M}'(\mathbf{y})) = c_i(x'_i, \mathcal{M}(\mathbf{x}')) - \min_{\mathbf{x}: x_i = y_i} c_i(x_i, \mathcal{M}(\mathbf{x})) \leq c$ . Instead,

$$\begin{aligned} c'_i(x'_i, \mathcal{M}'(\mathbf{y})) &= \max\{d(x'_i, f(\mathbf{y})), d(y_i, f(\mathbf{y}))\} + p_i(\mathbf{y}) = d(x'_i, f(\mathbf{y})) - p_i(\mathbf{y}) \\ &= d(x'_i, y_i) + c'_i(y_i, \mathcal{M}'(\mathbf{y})) = c + c'_i(y_i, \mathcal{M}'(\mathbf{y})). \end{aligned}$$

$\square$

**Lemma 9** *If there is  $\mathbf{y}$  with  $y_i = A_i - c$ ,  $c > 0$ , and  $c_i(y_i, \mathcal{M}(\mathbf{y})) < \mu_i + c$ , then there is another direct-revelation optimal OSP mechanism  $\mathcal{M}'$  that is less demanding to  $i$  than  $\mathcal{M}$ .*

*Similarly, if there is  $\mathbf{y}$  with  $y_i = B_i + c$  and  $c_i(y_i, \mathcal{M}(\mathbf{y})) < \mu_i + c$ , then there is another direct-revelation optimal OSP mechanism  $\mathcal{M}'$  that is less demanding to  $i$  than  $\mathcal{M}$ .*

**Proof** By Lemmas 4, 5, and 8, we know that  $c_i(y_i, \mathcal{M}(\mathbf{y})) = \mu_i$  for every  $\mathbf{y}$  such that  $y_i \in [A_i, B_i]$ .

Let  $y_i^A = \max_{c>0} \{y_i = A_i - c : c_i(y_i, \mathcal{M}(\mathbf{y})) < \mu_i + c\}$  and  $y_i^B = \min_{c>0} \{y_i = B_i + c : c_i(y_i, \mathcal{M}(\mathbf{y})) < \mu_i + c\}$ . Denote as  $y_i^*$  the one closer to the interval  $[A_i, B_i]$ , i.e.  $y_i^* = \arg \min_{y_i=y_i^A, y_i^B} \min\{d(A_i, y_i), d(B_i, y_i)\}$ . Henceforth, we assume w.l.o.g. that  $y_i^* = y_i^A$ . Let also  $\mathbf{y}^* = \min_{\hat{\mathbf{y}}: \hat{y}_i=y_i^*} c_i(\hat{y}_i, \mathcal{M}(\hat{\mathbf{y}}))$ . Observe that, by definition, it must be the case that  $c_i(y_i^*, \mathcal{M}(\mathbf{y}^*)) < \mu_i + c^*$ , where  $c^* = d(y_i^*, A_i)$ .

Consider  $\mathcal{M}' = (f, \mathbf{p}')$  as follows: for every  $\mathbf{y}$  such that  $y_i \in (A_i - c^*, B_i + c^*)$ , set payments such that  $c'_i(y_i, \mathcal{M}'(\mathbf{y})) = d(y_i, f(\mathbf{y})) + p'_i(\mathbf{y}) = c_i(y_i, \mathcal{M}(\mathbf{y})) - (\mu_i + c^* - c_i(y_i^*, \mathcal{M}(\mathbf{y}^*)))$ ; for every  $\mathbf{y}$  such that  $y_i \in \{y_i^*, B_i + c^*\}$ , set payments such that  $c'_i(y_i, \mathcal{M}'(\mathbf{y})) = c_i(y_i^*, \mathcal{M}(\mathbf{y}^*))$ ; for every remaining profile  $\mathbf{y}$ , set payments such that  $c'_i(y_i, \mathcal{M}'(\mathbf{y})) = c_i(y_i, \mathcal{M}(\mathbf{y}))$ .

Clearly,  $p'_i(\mathbf{y}) < p_i(\mathbf{y})$  for every  $\mathbf{y}$  such that  $y_i \in (A_i - c^*, B_i + c^*)$ , whereas  $p'_i(\mathbf{y}) \leq p_i(\mathbf{y})$  for every other profile  $\mathbf{y}$ . Moreover,  $\mathcal{M}'$  is a direct-revelation mechanism, and, since it places the facility in the median location, it is optimal if it is OSP. Hence, it is only left to show that  $\mathcal{M}'$  is OSP.

It is immediate to see that the OSP condition holds if from every profile one moves to another profile in which the location of  $i$  is  $x_i < y_i^*$  or  $x_i > B_i + c^*$ , and if from a profile in which the location of  $i$  is  $x_i \in (A_i - c^*, B_i + c^*)$  one moves to another profile in which the location of  $i$  is  $y_i \in (A_i - c^*, B_i + c^*)$ . Next we show that even if the real location of  $i$  is  $x_i \leq y_i^*$  or  $x_i \geq B_i + c^*$ , then it is not convenient for  $i$  to declare  $y_i \in (A_i - c^*, B_i + c^*)$ . That is, we prove that  $c'_i(x_i, \mathcal{M}'(\mathbf{x})) \leq c'_i(x_i, \mathcal{M}'(\mathbf{y}))$  for every  $\mathbf{x}$  and  $\mathbf{y}$  such that  $x_i$  and  $y_i$  are as above.

To this aim, let  $\mu'_i = c_i(y_i^*, \mathcal{M}(\mathbf{y}^*)) - c^*$ . By construction,  $c'_i(x_i, \mathcal{M}'(\mathbf{x}))\mu'_i$  when  $x_i \in [A_i, B_i]$ ,  $c'_i(x_i, \mathcal{M}'(\mathbf{x})) = \mu'_i + c$  when  $x_i = A_i - c$  or  $x_i = B_i + c$ , with  $0 \leq c \leq c^*$ , whereas  $c'_i(x_i, \mathcal{M}'(\mathbf{x})) \leq \mu'_i + c$  when  $x_i = A_i - c$  or  $x_i = B_i + c$ , with  $c > c^*$ . Then, the OSP condition can be proved as in (4) and (5) with  $\mu'_i$  in place of  $m_i$ .  $\square$

In conclusion, a direct-revelation mechanism  $\mathcal{M}$  that is OSP, optimal and for which the sum of payments assigned to  $i$  cannot be lowered must be exactly as OIM with  $\mu_i$  in place of  $m_i$ . However, suppose w.l.o.g. that  $m_i = R_i - A_i$  and consider a profile  $\mathbf{x}$  such that  $x_i = A_i$  and  $f(\mathbf{x}) = R_i$ . According to Lemma 2, such a profile surely exists. Moreover, as showed above,  $c_i(x_i, \mathcal{M}(\mathbf{x})) = \mu_i$ . Therefore, if all payments are non-positive, then

$$\mu_i = c_i(x_i, \mathcal{M}(\mathbf{x})) = d(x_i, f(\mathbf{x})) - p_i(\mathbf{x}) \geq d(x_i, f(\mathbf{x})) = R_i - A_i = m_i.$$

□

Theorem 6 does not rule out the existence of a *non-direct-revelation* mechanism that is OSP with lower payments. However, in the next section we show that for *every* mechanism there is at least one instance on which it cannot set payments lower than the one assigned by OIM.

#### 5.4 Payments in Non-direct-revelation Mechanisms can be High

We prove that for every mechanism there is at least one instance on which at least  $\left\lceil \frac{n}{2} \right\rceil - 1$  agents incur in a very large cost, namely  $b - a$ , just as happens in OIM.

**Lemma 10** *For every optimal OSP mechanism with monitoring there is an instance of the facility location problem for which the mechanism sets payments at least as high as OIM.*

**Proof** Let  $\mathcal{M}$  be an optimal OSP mechanism for facility location with payment policy  $\mathbf{p}$  and let  $\delta$  be an arbitrary constant. Let  $t_0$  be the first time step during the execution of mechanism  $\mathcal{M}$  in which an agent diverges on types  $a$  and  $b - \delta$ ; call  $i_0$  the corresponding deviating agent. Similarly, for  $j > 0$ , let  $t_j$  be the first time step greater than or equal to  $t_{j-1}$  in which an agent, called  $i_j$ , diverges on types  $a$  and  $b - \delta$  given that the first  $j - 1$  agents that diverge on types  $a$  and  $b - \delta$  behave as follows:  $i_0, \dots, i_\ell$  take actions according to  $b - \delta$  whilst  $i_{\ell+1}, \dots, i_{j-1}$  take actions according to  $a$ ,  $\ell = \min\{j - 1, \frac{n-1}{2}\}$ . Note that  $\mathcal{M}$  may ask to more than one agent to diverge on types  $a$  and  $b - \delta$  at the same time (this explain why we allow  $t_j$  to be equal to  $t_{j-1}$ ). Moreover, there may be agents  $i_k$  that never diverge on types  $a$  and  $b - \delta$  (for which we set  $t_k = \infty$ ). However, these are at most  $\frac{n-1}{2}$ , otherwise  $\mathcal{M}$  must give the same output on instance  $\mathbf{x} = (b - \delta, \dots, b - \delta)$  and on the instance  $\mathbf{y}$  such that  $y_j = b - \delta$  if  $i$  diverges on  $a$  and  $b - \delta$ , and  $y_j = a$  otherwise. But this contradicts the optimality of  $\mathcal{M}$ .

We now show that for the following instance the payments assigned by  $\mathcal{M}$  are at least as high as the payment assigned by OIM. We consider the instance  $\mathbf{x}$  according to which the real position of agent  $i_j$  is  $x_j = b - \delta$  if  $j \leq \frac{n+1}{2}$ , and  $x_j = a$ , otherwise. It is not hard to see that OIM on this instance assigns a zero payment to every agent whose real position is  $a$ , and, among agents with real position  $b - \delta$ , only to the last to be queried. It assigns instead a payment of  $b - a$  to every remaining agent.

We next show that for  $j \leq \frac{n-1}{2}$ , it must be the case that  $\mathcal{M}$  also sets  $p_j^{\mathcal{M}}(\mathbf{x}) \leq \delta - (b - a)$ . Let indeed  $t'$  be the first step within mechanism  $\mathcal{M}$  in which  $i_j$  diverges on types  $b - \delta$  and  $b$ , given that the first  $j - 1$  agents that diverge on types  $a$  and  $b - \delta$  take actions according to  $b - \delta$ , and the remaining agents take actions according to  $a$ . We set  $t^* = \min\{t', t_j\}$ . Consider then the following instance  $\mathbf{y}$ :  $y_j = b$ ,  $y_k = b - \delta$  if  $t_k < t^*$  and  $k \neq j$ , and  $y_k = a$  otherwise. Note that there are at most  $j - 1 \leq \frac{n-1}{2} - 1$  agents whose location is  $b - \delta$  and at least  $\frac{n+1}{2}$  agents whose

location is  $a$ . Then, by optimality of  $\mathcal{M}$ , we have that  $f(\mathbf{y}) = a$ . Moreover, since  $\mathcal{M}$  is OSP, it follows that

$$\begin{aligned} d(y_j, f(\mathbf{y})) - p_j^{\mathcal{M}}(\mathbf{y}) &\leq \max\{d(y_j, f(\mathbf{x})), d(x_j, f(\mathbf{x}))\} - p_j^{\mathcal{M}}(\mathbf{x}) \\ &\Rightarrow p_j^{\mathcal{M}}(\mathbf{x}) \leq \delta - (b - a) + p_j^{\mathcal{M}}(\mathbf{y}). \end{aligned}$$

Since  $p_j^{\mathcal{M}}(\mathbf{y}) \leq 0$ , the claim then follows by having  $\delta$  going to 0.  $\square$

The lemma above does not exclude that an indirect optimal mechanism might set payments to the agents smaller than OIM's only for some specific order in which agents are queried. It is left open to understand if this is the case. However, we remark that OIM maintains OSP irrespectively of such an ordering.

## 6 Conclusions

We have studied the limitations of OSP mechanisms in terms of the approximation guarantee of their outputs. By focusing on two paradigmatic problems in the literature, machine scheduling and facility location, we have shown that OSP can yield a significant loss in the quality of the solutions returned. We have proposed the use of a novel mechanism design paradigm, namely monitoring, as a way to reconcile OSP with good approximations. Our positive results show how the ingredients needed for truthfulness with monitoring marry up the demands needed for OSP.

We leave open the problem of understanding the extent to which this parallel holds in general. Several additional open problems pertain the two case studies considered. For machine scheduling, it would be interesting to see whether the lower bound can be improved or a matching upper bound can be proved. Recent follow-up works [18, 19] give some answers to this question for the special case of “small” domains. For facility location, it is interesting to establish if indirect mechanisms can be more frugal for the agents. More generally, the mechanisms with monitoring for which we provide an OSP implementation are shown to be collusion-resistant; is there any way to guarantee OSP (with monitoring) without relying so heavily on coalitional notions of incentive-compatibility? And how hard is it to design OSP mechanisms that do not use any additional control on agents' declarations? [18, 19] make some initial progress in these directions.

Finally, it would be interesting even to analyze whether the results in this work continue to hold with stronger and weaker variants of OSP, as the ones considered in [15, 38, 43].

**Acknowledgements** Diodato Ferraioli was supported by “GNCS-INdAM” and by the Italian MIUR PRIN 2017 Project ALGADIMAR “Algorithms, Games, and Digital Markets”. Carmine Ventre was supported by the EPSRC Grant EP/M018113/1.

**Funding** Open access funding provided by Università degli Studi di Salerno within the CRUI-CARE Agreement.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adamczyk, M., Borodin, A., Ferraioli, D., de Keijzer, B., Leonardi, S.: Sequential posted price mechanisms with correlated valuations. In: WINE 2015, pp. 1–15 (2015)
- Archer, A., Tardos, É.: Truthful mechanisms for one-parameter agents. In: FOCS 2001, pp. 482–491 (2001)
- Ashlagi, I., Gonczarowski, Y.A.: Stable matching mechanisms are not obviously strategy-proof. *J. Econ. Theory* **177**, 405–425 (2018)
- Ashlagi, I., Dobzinski, S., Lavi, R.: Optimal lower bounds for anonymous scheduling mechanisms. *Math. Oper. Res.* **37**(2), 244–258 (2012)
- Babaioff, M., Immorlica, N., Lucier, B., Weinberg, S.M.: A simple and approximately optimal mechanism for an additive buyer. In: FOCS 2014, pp. 21–30 (2014)
- Badem, S., Gonczarowski, Y.A.: Gibbard–Satterthwaite success stories and obvious strategyproofness. In: EC 2017, p. 565 (2017)
- Brânzei, S., Procaccia, A.D.: Verifiably truthful mechanisms. In: ITCS 2015, pp. 297–306 (2015)
- Caragiannis, I., Elkind, E., Szegedy, M., Lan, Y.: Mechanism design: from partial to probabilistic verification. In: EC 2012, pp. 266–283 (2012)
- Chawla, S., Hartline, J.D., Malec, D.L., Sivan, B.: Multi-parameter mechanism design and sequential posted pricing. In: STOC 2010, pp. 311–320 (2010)
- Christodoulou, G., Kovács, A.: A deterministic truthful PTAS for scheduling related machines. *SIAM J. Comput.* **42**(4), 1572–1595 (2013)
- Dobzinski, S., Vondrák, J.: The computational complexity of truthfulness in combinatorial auctions. In: EC 2012, pp. 405–422 (2012)
- Elkind, E., Sahai, A., Steiglitz, K.: Frugality in path auctions. In: SODA 2004, pp. 701–709 (2004)
- Ferraioli, D., Ventre, C.: Obvious strategyproofness needs monitoring for good approximations. In: AAAI 2017, pp. 516–522 (2017)
- Ferraioli, D., Ventre, C.: Probabilistic verification for obviously strategyproof mechanisms. In: IJCAI 2018 (2018)
- Ferraioli, D., Ventre, C.: Obvious strategyproofness, bounded rationality and approximation: the case of machine scheduling. In: SAGT 2019 (2019)
- Ferraioli, D., Ventre, C., Aranyi, G.: A mechanism design approach to measure awareness. In: AAAI 2015, pp. 886–892 (2015)
- Ferraioli, D., Serafino, P., Ventre, C.: What to verify for optimal truthful mechanisms without money. In: AAMAS 2016, pp. 68–76 (2016)
- Ferraioli, D., Meier, A., Penna, P., Ventre, C.: Obviously strategyproof mechanisms for machine scheduling. In: ESA 2019 (2019)
- Ferraioli, D., Meier, A., Penna, P., Ventre, C.: Automated optimal OSP mechanisms for set systems—the case of small domains. In: WINE 2019, pp. 171–185 (2019)
- Fotakis, D., Tzamos, C.: On the power of deterministic mechanisms for facility location games. *ACM Trans. Econ. Comput.* **2**(4), 15:1–15:37 (2014)
- Fotakis, D., Krysta, P., Ventre, C.: Combinatorial auctions without money. In: AAMAS 2014, pp. 1029–1036 (2014)
- Halldórsson, M.M.: Approximations of weighted independent set and hereditary subset problems. In: COCOON 1999, pp. 261–270 (1999)

23. Hochbaum, D.S., Shmoys, D.B.: A polynomial approximation scheme for scheduling on uniform processors: using the dual approximation approach. *SIAM J. Comput.* **17**(3), 539–551 (1988)
24. Hurkens, C.A.J., Schrijver, A.: On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems. *SIAM J. Discrete Math.* **2**(1), 68–72 (1989)
25. Karlin, A.R., Kempe, D.: Beyond vcg: frugality of truthful mechanisms. In: *FOCS 2005*, pp. 615–624 (2005)
26. Koutsoupias, E.: Scheduling without payments. *Theory Comput. Syst.* **54**(3), 375–387 (2014)
27. Koutsoupias, E., Vaidali, A.: A lower bound of  $1 + \varphi$  for truthful scheduling mechanisms. In: *MFCS 2007*, pp. 454–464 (2007)
28. Kovács, A., Meyer, U., Ventré, C.: Mechanisms with monitoring for truthful RAM allocation. In: *WINE 2015*, pp. 398–412 (2015)
29. Krysta, P., Telelis, O., Ventré, C.: Mechanisms for multi-unit combinatorial auctions with a few distinct goods. *J. Artif. Intell. Res.* **53**, 721–744 (2015)
30. Kyropoulou, M., Ventré, C.: Obviously strategyproof mechanisms without money for scheduling. In: *AAMAS 2019* (2019)
31. Lehmann, D., O’callaghan, L.I., Shoham, Y.: Truth revelation in approximately efficient combinatorial auctions. *J. ACM* **49**(5), 577–602 (2002)
32. Li, S.: Obviously strategy-proof mechanisms. *Am. Econ. Rev.* **107**(11), 3257–87 (2017)
33. Mackenzie, A.: A revelation principle for obviously strategy-proof implementation (2017)
34. Moulin, H.: On strategy-proofness and single-peakedness. *Public Choice* **35**, 437–455 (1980)
35. Nisan, N., Ronen, A.: Algorithmic mechanism design. *Games Econ. Behav.* **35**, 166–196 (2001)
36. Penna, P., Ventré, C.: Optimal collusion-resistant mechanisms with verification. *Games Econ. Behav.* **86**, 491–509 (2014)
37. Procaccia, A.D., Tennenholtz, M.: Approximate mechanism design without money. *ACM Trans. Econ. Comput.* **1**(4), 18:1–18:26 (2013)
38. Pycia, M., Troyan, P.: Obvious dominance and random priority. In: *EC 2019* (2019)
39. Sandholm, T., Gilpin, A.: Sequences of take-it-or-leave-it offers: near-optimal auctions without full valuation revelation. In: *AMET 2003*, pp. 73–91 (2003)
40. Serafino, P., Ventré, C., Vaidali, A.: Truthfulness on a budget: trading money for approximation through monitoring. In: *AAMAS 2019*, pp. 1234–1242 (2019)
41. Talwar, K.: The price of truth: frugality in truthful mechanisms. In: *STACS 2003*, pp. 608–619 (2003)
42. Ventré, C.: Truthful optimization using mechanisms with verification. *Theor. Comput. Sci.* **518**, 64–79 (2014)
43. Zhang, L., Levin, D.: Partition obvious preference and mechanism design: theory and experiment. *Am. Econ. Rev.* **107**(5), 235–39 (2017)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.